

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2005, 04:00:28 ; Search time 3259.53 Seconds

(without alignments)

2020.266 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVTIQHFWKRTLGPFYPS.....HAERAIPVSRREEKTPSAPS 173

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10657740/runat_27052005_165253_3329/app_query.fasta_1.590
-DB=EST -QFWT=fascap -SURFEX=first -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=150
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10657740 @CGN 1.1 5533 @runat_27052005_165253_3329 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gest1:*

9: gb_gest2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	100.0	522	9 AY419529	Homo sapi
2	916	100.0	536	2 BF726399	BF726399 by06a09.y
3	916	100.0	576	2 BF726236	BF726236 by03f01.y
4	916	100.0	578	4 BM721893	BM721893 UI-E-E00-
5	916	100.0	592	2 BF727295	BF727295 by19e10.y
6	916	100.0	592	4 BM705926	BM705926 UI-E-DW0-
7	916	100.0	596	4 BM696581	BM696581 UI-E-DW0-
8	916	100.0	607	6 CD675250	CD675250 fs21c02.y
9	916	100.0	629	2 BF727028	BF727028 by15g05.y

10	916	100.0	630	4 BM696651	BM696651 UI-E-DW0-
11	916	100.0	633	6 CD672144	CD672144 f910b07.y
12	916	100.0	659	4 BM706270	BM706270 UI-E-DW0-
13	916	100.0	681	5 BX118596	BX118596 BX118596
14	916	100.0	698	2 BF727324	BF727324 by19h12.y
15	916	100.0	724	4 BM722336	BM722336 UI-E-E00-
16	912	99.6	559	4 BM722779	BM722779 UI-E-E00-
17	911	99.5	635	4 BM697066	BM697066 UI-E-DW0-
18	910	99.3	532	2 BF726358	BF726358 by05d12.y
19	909	99.2	522	9 AY419530	AY419530 Pan trogl
20	907	99.0	577	2 BF726253	BF726253 by03h05.y
21	907	99.0	587	2 BF726422	BF726422 by06d05.y
22	906	98.9	597	2 BF726330	BF726330 by05b01.y
23	903	98.6	580	4 BM706139	BM706139 UI-E-DW0-
24	903	98.6	586	4 BM722650	BM722650 UI-E-E00-
25	901	98.4	569	4 BM697368	BM697368 UI-E-DW0-
26	901	98.4	661	4 BM697160	BM697160 UI-E-DW0-
27	900	98.3	604	5 BQ640267	BQ640267 he26c10.y
28	898	98.0	577	2 BF727002	BF727002 by15c11.y
29	896	97.8	518	2 BF726854	BF726854 by13a10.y
30	893	97.5	523	2 BF726438	BF726438 by06f01.y
31	893	97.5	588	4 BM696799	BM696799 UI-E-DW0-
32	892	97.4	593	4 BM697101	BM697101 UI-E-DW0-
33	889	97.1	558	2 BF726679	BF726679 by10d02.y
34	888	96.9	558	4 BM696667	BM696667 UI-E-DW0-
35	887	96.8	508	4 BM696489	BM696489 UI-E-DW0-
36	887	96.8	663	4 BM686748	BM686748 UI-E-CQ0-
37	880	96.1	503	2 BF727205	BF727205 by18b11.y
38	879	96.0	519	4 BM696477	BM696477 UI-E-DW0-
39	877	95.7	536	4 BM686206	BM686206 UI-E-C11-
40	876	95.6	542	5 BQ637175	BQ637175 he06f05.y
41	873	95.3	498	4 BM696853	BM696853 UI-E-DW0-
42	873	95.3	688	7 CF732528	CF732528 UI-M-HA0-
43	871	95.1	522	9 AY419531	AY419531 Mus muscu
44	871	95.1	552	7 CK628252	CK628252 ip09h08.y
45	871	95.1	557	7 CK627503	CK627503 io09g04.y
46	871	95.1	558	7 CK627631	CK627631 io06f03.y
47	871	95.1	566	7 CK627600	CK627600 io06c02.y
48	871	95.1	570	7 CK628219	CK628219 ip09d07.y
49	871	95.1	574	7 CK627249	CK627249 io01b06.y
50	871	95.1	579	7 CK628200	CK628200 ip09b11.y
51	871	95.1	582	7 CK628246	CK628246 ip09g11.y
52	871	95.1	582	7 CK628391	CK628391 ip11f11.y
53	871	95.1	586	7 CK627945	CK627945 ip04h08.y
54	871	95.1	590	7 CK627884	CK627884 ip04a08.y
55	871	95.1	590	7 CK628336	CK628336 ip10h08.y
56	871	95.1	590	7 CK628526	CK628526 ip13f05.y
57	871	95.1	594	7 CK628372	CK628372 ip11d12.y
58	871	95.1	601	7 CK627483	CK627483 io04d12.y
59	871	95.1	601	7 CK627500	CK627500 io04f12.y
60	871	95.1	606	6 CB849766	CB849766 MRA-1515
61	871	95.1	606	7 CK628546	CK628546 ip13b12.y
62	871	95.1	615	7 CK627740	CK627740 ip02b06.y
63	871	95.1	617	6 CB848065	CB848065 M2PN-3875
64	871	95.1	621	6 CB845999	CB845999 M2PN-1715
65	871	95.1	622	7 CK627736	CK627736 ip02b02.y
66	871	95.1	623	7 CK627777	CK627777 ip02f06.y
67	871	95.1	626	7 CK627695	CK627695 ip01e09.y
68	871	95.1	626	7 CK627737	CK627737 ip02b03.y
69	871	95.1	629	6 CB055541	CB055541 NISC_j106
70	871	95.1	633	7 CK627814	CK627814 ip03b02.y
71	871	95.1	636	7 CK627616	CK627616 io06d08.y
72	871	95.1	637	7 CK627374	CK627374 io02h07.y
73	871	95.1	642	7 CK627318	CK627318 io02b04.y
74	871	95.1	643	6 CB846376	CB846376 M2PN-2110
75	871	95.1	643	7 CK627445	CK627445 io03h09.y
76	871	95.1	645	6 CB848071	CB848071 M2PN-3881
77	871	95.1	645	7 CK627651	CK627651 io06h10.y
78	871	95.1	662	6 CB841336	CB841336 M158-1855
79	871	95.1	665	6 CB849833	CB849833 MRA-1582
80	871	95.1	669	6 CB840876	CB840876 M158-1296
81	871	95.1	676	6 CB840784	CB840784 M158-1183
82	871	95.1	690	6 CB846475	CB846475 M2PN-2213

83	871	95.1	691	7	CF731811	UI-M-HA0-	CF731811	AY419529	LOCUS	522 bp	DNA	linear	GSS 17-DEC-2000
84	871	95.1	696	7	CO429233	UI-M-HW0-	CO429233	AY419529	DEFINITION	522 bp	DNA	linear	GSS 17-DEC-2000
85	871	95.1	699	7	CO427334	UI-M-HW0-	CO427334	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
86	871	95.1	705	7	CO429948	UI-M-HW0-	CO429948	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
87	871	95.1	709	7	CO428729	UI-M-HW0-	CO428729	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
88	871	95.1	721	7	CF733985	UI-M-HA0-	CF733985	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
89	871	95.1	851	6	CB845124	M2PN-0694	CB845124	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
90	869	94.9	677	7	CB435619	BE030013B	CB435619	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
91	867	94.7	600	4	BG805475	0693-23 M	BG805475	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
92	867	94.7	834	6	CB845009	M2PN-0575	CB845009	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
93	866	94.5	527	2	BF726880	by13d10.Y	BF726880	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
94	866	94.5	536	2	BF726381	by05911.Y	BF726381	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
95	866	94.5	600	6	CA527708	8051-64 M	CA527708	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
96	866	94.5	637	6	CB848191	M2PN-4009	CB848191	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
97	866	94.5	854	6	CB848176	MRA-0308	CB848176	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
98	865	94.4	619	7	CK627238	io01a02.Y	CK627238	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
99	865	94.4	625	7	CK627598	io06b10.Y	CK627598	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
100	865	94.4	625	7	CK629950	UI-M-HW0-	CK629950	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
101	865	94.4	711	7	CK429062	UI-M-HW0-	CK429062	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
102	863	94.2	572	7	CK627552	io05e04.Y	CK627552	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
103	863	94.2	600	4	BG808732	2121-59 M	BG808732	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
104	862	94.1	745	7	CN441369	BE04025B2	CN441369	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
105	862	94.1	809	7	CN438019	BE04014A2	CN438019	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
106	860	93.9	556	6	CB848450	M2PN-4276	CB848450	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
107	858	93.7	706	6	CB842697	M15B-3361	CB842697	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
108	857	93.6	645	6	CB842016	M15B-2649	CB842016	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
109	856	93.4	581	7	CF732553	UI-M-HA0-	CF732553	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
110	856	93.4	581	7	CO429470	UI-M-HW0-	CO429470	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
111	856	93.4	601	7	CO429510	UI-M-HW0-	CO429510	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
112	856	93.4	705	7	CO429072	UI-M-HW0-	CO429072	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
113	856	93.4	760	7	CO427914	UI-M-HW0-	CO427914	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
114	856	93.4	771	7	CO427433	UI-M-HW0-	CO427433	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
115	856	93.4	773	7	CO428121	UI-M-HW0-	CO428121	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
116	856	93.4	897	7	CR550671	CR550671	CR550671	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
117	854	93.2	506	2	BF727221	by18e03.Y	BF727221	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
118	852	93.0	654	6	CB841950	M15B-2579	CB841950	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
119	851	92.9	653	6	CB850107	MRA-1868	CB850107	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
120	851	92.9	798	7	CO427114	UI-M-HW0-	CO427114	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
121	850	92.8	645	7	CO427082	UI-M-HW0-	CO427082	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
122	850	92.8	698	7	CN457234	UI-M-HF0-	CN457234	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
123	850	92.8	700	7	CO427294	UI-M-HW0-	CO427294	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
124	849.5	92.7	746	7	CF731540	UI-M-HA0-	CF731540	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
125	849.5	92.7	605	7	CK627795	ip02h03.Y	CK627795	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
126	849.5	92.7	633	7	CO427584	UI-M-HW0-	CO427584	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
127	849.5	92.7	813	6	CB848822	MRA-0355	CB848822	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
128	849	92.7	485	2	BF727264	by19b05.Y	BF727264	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
129	848	92.6	552	7	CKG28267	ip10b01.Y	CKG28267	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
130	847	92.5	478	4	BM705752	UI-E-DW0-	BM705752	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
131	847	92.5	698	7	CO429320	UI-M-HW0-	CO429320	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
132	846	92.4	524	2	BF726441	by06f04.Y	BF726441	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
133	845	92.2	539	7	CF732589	UI-M-HA0-	CF732589	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
134	845	92.2	580	7	CF731803	UI-M-HA0-	CF731803	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
135	845	92.2	650	6	CB842520	M15B-3174	CB842520	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
136	845	92.2	654	6	CB842155	M15B-2792	CB842155	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
137	845	92.2	769	7	CF732285	UI-M-HA0-	CF732285	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
138	844.5	92.2	778	7	CO427386	UI-M-HW0-	CO427386	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
139	844	92.1	759	7	CF731519	UI-M-HA0-	CF731519	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
140	842	91.9	522	2	BF726109	by01b12.Y	BF726109	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
141	839	91.6	597	7	CF732732	UI-M-HA0-	CF732732	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
142	838	91.5	479	2	BF727402	by20h11.Y	BF727402	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
143	838	91.5	519	2	BF726860	by13b07.Y	BF726860	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
144	837	91.4	473	2	BF726931	by14c03.Y	BF726931	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
145	837	91.4	585	6	CB848256	M2PN-4074	CB848256	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
146	837	91.4	642	4	BM721028	UI-E-E00-	BM721028	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
147	836.5	91.3	686	7	CO427562	UI-M-HW0-	CO427562	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
148	835	91.2	516	2	BF727010	by15d10.Y	BF727010	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
149	835	91.2	522	2	BF726595	by09a09.Y	BF726595	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000
150	835	91.2	560	7	CK627898	ip04c07.Y	CK627898	AY419529	Genomic survey sequence.	522 bp	DNA	linear	GSS 17-DEC-2000

ALIGNMENTS

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

BY"
 /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
 from different adults (both approximately 40 years old)
 together yielded 20ug of total RNA and 150ng mRNA for cDNA
 library synthesis. A directionally cloned cDNA library in
 the pCMVSPORT6 vector was constructed at Life
 Technologies, essentially following the protocols of the
 SuperScript Plasmid System full details of which are
 contained in the manufacturer's instruction manual
 (http://www.lifetech.com/). First strand synthesis was
 carried out using a Not I primer-adaptor
 [5'-pGATAGTTCAGATCGAGCGGCCCT(T)15-3']. Not I/blunt
 end inserts were cloned into the Not I/EcoR V sites in the
 vector. EST analysis was performed on the unamplified
 library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
 Pred. No.: 6,17e-101 Length: 576
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF726236 (1-576)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
 Db 48 ATGACGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCAGC 107
 Qy 21 ArgLeuPheAspGlnPhePheGlyGlyLeuPheGlyTrpAspLeuLeuProPheLeu 40
 Db 108 CGCTGTTTCGACCATCTTTCGGCGAGGGCCCTTTTGAGTATGACCTCTGCCCTTCTCTG 167
 Qy 41 SerSerThrIleSerProTyTrpArgGlnSerLeuPheArgThrValLeuAspSergly 60
 Db 168 TCGTCCACCATCAGCCCTACTACCGCAGCTCCCTTTCGACCGTCTGGATCCGGC 227
 Qy 61 IleSerGluValArgSerAspArgPheValIlePheLeuAspValIleHisPhe 80
 Db 228 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTCTGTCATCTTCTCGATGTGAAGCACTC 287
 Qy 81 SerProGluAspLeuThrValIleValGlnAspPheValGluIleHisGlyIleHis 100
 Db 288 TCCCGGAGGACCTCACCCTGAAGTGCAGGACGACCTTTGTGGAGATCCAGGAAGCAC 347
 Qy 101 AsnGluArgGlnAspAspHisGlyTyTrpIleSerArgGluPheHisArgArgTyArgLeu 120
 Db 348 AACGAGCGCCAGGACGACCATCTTCCGTGAGTTCACCGCGCTACCGCTG 407
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 408 CCGTCCAACTGGACGACGATCGGCCCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 467
 Qy 141 PheCysGlyProLysIleGlnThrClyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 468 TTCTGTGGCCCAAGATCCAGACTGGCTGTGATGCCACCCAGCGCGAGGACCATCCCC 527
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 528 GTGTCCGGGAGGAGAGACCCACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 566

RESULT 4

BM721893
 LOCUS 578 bp mRNA linear EST 01-MAR-2002
 DEFINITION UI-E-E00-ahw-c-15-0-UI r1 UI-E-E00 Homo sapiens cDNA clone
 UI-E-E00-ahw-c-15-0-UI 5', mRNA sequence.
 ACCESSION BM721893
 VERSION BM721893.1 GI:19042101
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE

AUTHORS
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

GENOME RES. 6 (9), 791-806 (1996)

MEDLINE

97044477

COMMENT

CONTACT: Soares, MB

COORDINATED LABORATORY FOR COMPUTATIONAL GENOMICS

UNIVERSITY OF IOWA

375 NEWTON ROAD

4156 MEBRF, IOWA CITY, IA 52242, USA

TEL: 319 335 8250

FAX: 319 335 9565

EMAIL: bento-soares@uiowa.edu

TISSUE PROCUREMENT: Dr. Gregg Hageman

CDNA LIBRARY PREPARATION: Dr. M. Bento Soares, University of Iowa

CDNA LIBRARY ARRAYED BY: Dr. M. Bento Soares, University of Iowa

DNA SEQUENCING BY: Dr. M. Bento Soares, University of Iowa

CLONE DISTRIBUTION: Researchers may obtain clones from Research

GENETICS (WWW.RESGEN.COM).

SEQ PRIMER: M13 Reverse.

FEATURES

Location/Qualifiers

1..578

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="UI-E-E00-ahw-c-15-0-UI"

/tissue_type="fetal eye"

/dev_stage="fetal"

/lab_hosts="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-E00"

/notes="Organ: eye; Vector: pTTT3-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-E-E00 is a cDNA library containing the following

tissue(s): fetal eye. The library was constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into pTTT3-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CGGATACC. This library

was created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 6.2e-101 Length: 578
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM721893 (1-578)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
 Db 55 ATGACGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTTGGGGCCCTTCTACCCAGC 114
 Qy 21 ArgLeuPheAspGlnPhePheGlyGlyLeuPheGlyTrpAspLeuLeuProPheLeu 40
 Db 115 CGCTGTTTCGACCATCTTTCGGCGAGGGCCCTTTTGAGTATGACCTCTGCCCTTCTCTG 174
 Qy 41 SerSerThrIleSerProTyTrpArgGlnSerLeuPheArgThrValLeuAspSergly 60
 Db 175 TCGTCCACCATCAGCCCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 234
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPhe 80

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 578)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road

4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

```

Db      235 ATCTCTGAGGTTTCGATCCGACCGGACAGTTTCGTCATCTTCTCGATGTGAAGCACATTC 294
Qy      81 SerProGluAspLeuThrValGlnAspAspPheValGluIleHisGlyLysHis 100
Db      295 TCCCGGAGGACCTCACCGTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAAGCAC 354
Qy      101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
Db      355 AACGAGCGCCAGGACGACCGCTACATTTCCCGTGAGTTCCACCGCGGTACCGCGCTG 414
Qy      121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db      415 CCGTCCACGTTGACAGTGGCGCCCTCTTGTCTCCCTGCTGCGATGGCATGTGAC 474
Qy      141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db      475 TTCTGTGCGCCCAAGATCCAGACTGGCTGGATGCCACCGCCGCGAGCGACCATCCCC 534
Qy      161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db      535 GTGTGCGGGAGGAGAGACCCACCTCGGCTCCCTCGTCC 573

RESULT 5
BF727295 592 bp mRNA linear EST 05-JAN-2001
LOCUS by19e10.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
DEFINITION sapiens cDNA clone by19e10 5', mRNA sequence.
ACCESSION BF727295
VERSION BF727295.1 GI:12043206
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 592)
JOURNAL Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
COMMENT Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 19 row: e column: 10
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1..592
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="by19e10"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
By"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life
Technologies, essentially following the protocols of the
Superscript Plasmid System full details of which are
contained in the manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTCTAGATCCGAGCGGCCCT(T)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."
```

```

ORIGIN
Alignment Scores:
Pred. No.: 6.41e-101 Length: 592
Score: 916.00 Matches: 173
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF727295 (1-592)
Qy      1 MetAspValThrIleGlnHisProThrPheLysArgThrLeuGlyProPheTyrProSer 20
Db      11 ATGGATGTGACATCCAGCACCCCTGGTTCAAGCGCACCTTGGGGCCCTTCTACCCACG 70
Qy      21 ArgLeuPheAspGlnPhePheGlyGluLysLeuPheGluTyrAspLeuLeuProPheLeu 40
Db      71 CGGCTGTTTCGACCAAGTTTTTTCGGCGAGGCGCTTTTGTAGTATGACCTGCTGCCCTTC 130
Qy      41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db      131 TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTTTCCGACCGTCTCGACTCCGGC 190
Qy      61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db      191 ATCTCTGAGGTTTCGATCCGACCGGACAAAGTTCGTCTATCTTCTCGATGTGAAGCACTTC 250
Qy      81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
Db      251 TCCCGGAGGACCTCACCGTGAAGTGCAGGACACTTTGTGGAGATCCACGGAAAGCAC 310
Qy      101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
Db      311 AACGAGCGCCAGGACGACCGGCTACATTTCCGTTGAGTTCACCGCGGCTACCGCGCTG 370
Qy      121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db      371 CCGTCCAACTGGACCACTGCGGCCCTCTCTTGTCTCCCTGTCTGCGATGTCGTGACC 430
Qy      141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db      431 TTCTGTGCGCCCAAGATCCAGACTGGCTGGATGCCACCGCCGCGAGCGACCATCCCC 490
Qy      161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db      491 GTGTGCGGGAGGAGAGACCCACCTCGGCTCCCTCGTCC 529

RESULT 6
BF705926 592 bp mRNA linear EST 28-FEB-2002
LOCUS UI-E-DW0-agh-f-18-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
DEFINITION UI-E-DW0-agh-f-18-0-UI 5', mRNA sequence.
ACCESSION BF705926
VERSION BF705926.1 GI:19019184
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 592)
JOURNAL Bonaldo,M.F., Lennon,G. and Soares,M.B.
COMMENT Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
```

Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

source

1..592
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agh-f-18-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pTT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 6,41e-101 Length: 592
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM705926 (1-592)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
 Db 52 ATGGACGTGACCATCCAGCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCAGC 111
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluLeuPheGluTyAspLeuLeuProPheLeu 40
 Db 112 CGCTGTTCGACAGTTTTCGGCAGGGCCCTTTTGTAGTAGTACCTGCTGCCCTTCCTG 171
 Qy 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 172 TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCGACCGTCTCGACTCCGCG 231
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 232 ATCTCTGAGTTTCGATCCGACGGGACAAAGTTTCGTTCATCTTCCTCGATGTGAAGCACTTC 291
 Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 292 TCCCCGGAGGACCTCACCCTGAGGTGAGGACGACGCTTTTGGGATCCACGGAAAGCAC 351
 Qy 101 AsnGluArgGlnAspAspHisGlyTyriLeSerArgGluPheHisArgArgTyArgLeu 120
 Db 352 AACGAGCGCCAGGACACCGGCTACATTTCCGTGAGTTCCACCGCGCTACCGCTG 411
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 412 CCGTCCAACTGGAGCAGTCGCGCCCTCTCTTGTCTCCCTGCTGCGGATGGCATGCTGACC 471

Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 472 TTCTGTGGCCCCAAGATCCAGACTGGCTGATGCCACCCAGCGCGAGGAGCATCCCC 531
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 532 GTGTGCGGGAGGAGAGAGCCACCTCGGCTCCCTCGTCC 570

RESULT 7

BM696581

LOCUS

DEFINITION

UI-E-DW0-agk-o-15-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone

UI-E-DW0-agk-o-15-0-UI 5', mRNA sequence.

ACCESSION

BM696581

VERSION

BM696581.1

GI:19009839

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 596)

AUTHORS

Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

8889548

COMMENT

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..596

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-DW0-agk-o-15-0-UI"

/tissue_type="lens"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-DW0"

/note="Organ: eye; Vector: pTT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pTT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 6,47e-101 Length: 596
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-10-657-740-1 (1-173) x BM696581 (1-596)

QY 1 MetAspValThrIleGlnHisProTTPheLysArgThrLeuGlyProPheTyrProSer 20
DB 56 ATGGATGTGACCATCAGACACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCCGAC 115
QY 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
DB 116 CGGCTGTTTCGACAGTTTTCGGCGAGGGCCCTTTTGGATGATGACCTGCTGCCCTTCCTG 175
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
DB 176 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTCCGACCGCTGCTGGACTCCGCG 235
QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallyHisPhe 80
DB 236 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTTCGTCATCTTCTCGATGGAAGCACTTC 295
QY 81 SerProGluAspLeuThrVallyValGlnAspAspPheValGluIleHisGlyLysHis 100
DB 296 TCCCGGAGGACCTCACCCTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 355
QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
DB 356 AACGAGCGCCAGGACGACCCAGCTTCCCTGAGTTCACCGCGCTTACCGCCTG 415
QY 121 ProSerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
DB 416 CGGTCCAAAGTCGACGACGTCGCGCCCTCTCTCTCCCTGCTGCGATGCACTGTCGACC 475
QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
DB 476 TTCTGTGCCCCAAGATCCAGACTGGCTGGATGTCACCCAGCGCGAGCGGACCATCCCC 535
QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
DB 536 GTGTCCGCGAGGAGAGCCACCTCGGCTCCCTCGTCC 574

RESULT 8
CD675250 607 bp mRNA linear EST 24-JUN-2003
LOCUS fs21c02.y1 Human Lens cDNA (Normalized): fs Homo sapiens cDNA clone
DEFINITION fs21c02.5', mRNA sequence.
ACCESSION CD675250
VERSION CD675250.1 GI:32176981
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 607)
AUTHORS Wistow,G., Bernstein,S.D., Wyatt,M.K., Behal,A., Touchman,J.W.,
Bouffard,G., Smith,D., and Peterson,K.
TITLE Expressed sequence tag analysis of adult human lens for the NEIBank
Project: over 2000 non-redundant transcripts, novel genes and
splice variants
JOURNAL Mol. Vis. 8 (4), 171-184 (2002)
MEDLINE 22103463
PUBMED 12107413
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 21 row: c column: 02
Seq primer: M13RF1 reverse primer (ABI).
LOCATION/Qualifiers
1. .607
FEATURES
source

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fs21c02"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Lens cDNA (Normalized): fs"
/library="Organ: Eye; Vector: pCMVSPORT6; A human lens
library (by) was normalized by self-subtraction. One
portion of double stranded plasmid DNA representing the
library was linearized by NotI. This NotI digested library
was used as a template for biotinylated RNA synthesis
using Sp6 RNA polymerase. Another portion of the double
stranded plasmid library was converted to single-stranded
circles in vitro using Gene II and Exonuclease III (Life
Technologies). Single-stranded DNA (1 mg) was hybridized
(Cot 500) with 41 mg of Bio-RNA and vector blocking
oligonucleotides. The hybridized Bio-RNA/ss-circles were
removed by streptavidin:phenol extraction. EST analysis
was performed on the library at the NIH Intramural
Sequencing Center (NISC)."
```

ORIGIN

Alignment Scores:

Pred. No.:	6.63e-101	Length:	607
Score:	916.00	Matches:	173
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-657-740-1 (1-173) x CD675250 (1-607)

QY	1	MetAspValThrIleGlnHisProTTPheLysArgThrLeuGlyProPheTyrProSer	20
DB	51	ATGGATGTGACCATCAGACACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCCGAC	110
QY	21	ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu	40
DB	111	CGGCTGTTTCGACAGTTTTCGGCGAGGGCCCTTTTGGATGATGACCTGCTGCCCTTCCTG	170
QY	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60
DB	171	TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCGACCGCTGCTGGACTCCGCG	230
QY	61	IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallyHisPhe	80
DB	231	ATCTCTGAGTTTCGATCCGACCGGACAAAGTTTCGTCATCTTCTCGATGGAAGCACTTC	290
QY	81	SerProGluAspLeuThrVallyValGlnAspAspPheValGluIleHisGlyLysHis	100
DB	291	TCCCGGAGGACCTCACCCTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC	350
QY	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
DB	351	AACGAGCGCCAGGACGACCCAGCTTCCCTGAGTTCACCGCGCTTACCGCCTG	410
QY	121	ProSerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	140
DB	411	CGGTCCAAAGTCGACGACGTCGCGCCCTCTCTTCTGCTCCCTGCTGCGATGCACTGTCGACC	470
QY	141	PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro	160
DB	471	TTCTGTGCCCCAAGATCCAGACTGGCTGGATGTCACCCAGCGCGAGCGGACCATCCCC	530
QY	161	ValSerArgGluGluLysProThrSerAlaProSerSer	173
DB	531	GTGTCCGCGAGGAGAGCCACCTCGGCTCCCTCGTCC	569

RESULT 9

BF727028 629 bp mRNA linear EST 05-JAN-2001

LOCUS BF727028

```

DEFINITION      by15g05.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
                  sapiens cDNA clone by15g05 5', mRNA sequence.
ACCESSION       BF727028
VERSION         BF727028.1  GI:12042939
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 629)
AUTHORS         Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
TITLE           NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL         Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT         Contact: Wistow G
                  Section on Molecular Structure and Function
                  National Eye Institute
                  6/331, NIH, Bethesda, MD 20892-2740, USA
                  Tel: 301 402 3452
                  Fax: 301 496 0078
                  Email: graeme@helix.nih.gov
                  Plate: 15 row: g column: 05
                  Seq primer: MJ3RP1 reverse primer (ABI).
FEATURES        Location/Qualifiers
                  1..629
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="by15g05"
                     /tissue_type="Lens"
                     /dev_stage="Adult"
                     /lab_host="EMDH10B"
                     /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                     BY"
                     /note="Organ: Eye; Vector: pCWSPORT6; Two human lenses
                     from different adults (both approximately 40 years old)
                     together yielded 20ug of total RNA and 150ng mRNA for cDNA
                     library synthesis. A directionally cloned cDNA library in
                     the pCWSPORT6 vector was constructed at Life
                     Technologies, essentially following the protocols of the
                     Superscript Plasmid System full details of which are
                     contained in the manufacturer's instruction manual
                     (http://www.lifetech.com/). First strand synthesis was
                     carried out using a Not I primer-adaptor
                     [5'-pGATGAGTTCTAGATCGGAGCGGCC(T)15-3']. Not I/blunt
                     end inserts were cloned into the Not I/EcoR V sites in the
                     vector. EST analysis was performed on the unamplified
                     library at the NIH Intramural Sequencing Center (NISC)."
```

Alignment Scores:

Pred. No.:	6,96e-101	Length:	629
Score:	916.00	Matches:	173
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-10-657-740-1 (1-173) x BF727028 (1-629)

QY	1	MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer	20
DB	48	ATGATGTGACCATCCAGCCCTGTTCAAGCGCACCTTGGGGCCCTTCTACCCAGC	107
QY	21	ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu	40
DB	108	CGGCTGTCGACAGTTTTCGCGAGGGCCCTTTTGTAGTATGACCTGCTGCGCTTCCTG	167
QY	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60
DB	168	TGTTCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCGACCGCTGTCATCCGAC	227
QY	61	IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPhe	80

DB	228	ATCTCTGAGGTTTCGATCCGACCGGACAAAGTTTCGTATCTTCTCTCGATGTGAAGCACTTC	287
QY	81	SerProGluAspLeuThrValIleValGlnAspAspPheValGluIleHisGlyLysHis	100
DB	288	TCCCGGAGGACCTCACCCTGAAGGTGAGGACGACTTTTGTGGAGATCCCGGAAGCAC	347
QY	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
DB	348	AACGAGCCGAGGACGACGACCGCTACATTTCCCTGAGTTCACCGCGCTACCGCCTG	407
QY	121	ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	140
DB	408	CGGTCCAAACGTGACGACGTCGCGCCCTCTCTGCTCCCTGTCGCGATGCTGACG	467
QY	141	PheCysGlyProLeuValIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro	160
DB	468	TTCGTGTCGCCCCAAGATCCAGACTGCGCTGGATGTCACCGCCGAGGAGCCATCCCC	527
QY	161	ValSerArgGluGluGlyProThrSerAlaProSerSer	173
DB	528	GTGTGCGGAGGAGAGACCCACCTCGGCTCCCTCGTCC	566

BM696651 630 bp mRNA linear EST 28-FEB-2002
 UI-E-DW0-agk-k-24-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 UI-E-DW0-agk-k-24-0-UI 5', mRNA sequence.

BM696651
 BM696651.1 GI:19009909
 EST.
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..630
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agk-k-24-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Ronaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dt)18 tail. The sequence tag for this library is CGATTAGCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:
 Pred. No.: 6,98e-101 Length: 630
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM696651 (1-630)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 54 ATGGATGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTCGGGCCCTTCTACCCACG 113
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 114 CGGCTGTTCGACCACTTTTTCGGCCAGGGCCCTTTTGTGATGATGACCTGCTGCCCTTCCTG 173
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 174 TCGTCCACATCAGCCCTTACTACCGCAGTCCCTTTCGGACCGTGTGACCTCCGGC 233
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValHisPhe 80
 Db 234 ATCTCTGAGTTTCGATCCAGCCGACAAAGTTTCGTCATCTTCTCGATGTGAAGCACTTC 293
 Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 294 TCCCGGAGGACCTCACCCTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 353
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 Db 354 AACGAGCCGAGGACGACGACGCTACATTTCCCGTGTGATTCACCGCGCTACCGCTG 413
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 414 CGTCCAACTGGACCACTGCGCCCTCTTCTGCTCCCTGCTGCGATGCGATGCTGACC 473
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 474 TTCTGTGCCCCAAGATCCAGACTGGCTGATGTCACCCACGCGGAGGAGCCATCCCC 533
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 534 GTGTCCGGAGGAGAGACCCACCTTCGGCTCCCTCGTCC 572

RESULT 11
 LOCUS CD672144
 DEFINITION fgl0b07.y2 Human Iris cDNA (Normalized): fg Homo sapiens cDNA clone
 fgl0b07.5', mRNA sequence.
 ACCESSION CD672144
 VERSION CD672144.1 GI:32173875
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 633)
 AUTHORS Wistow,G., Bernstein,S.L., Ray,S., Wyatt,M.K., Behal,A.,
 Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
 TITLE Expressed sequence tag analysis of adult human iris for the NEIBank
 Project: steroid-response factors and similarities with retinal
 pigment epithelium

JOURNAL Mol. Vis. 8 (4), 185-195 (2002)
 MEDLINE 22103462
 PUBMED 12107412
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 10 row: b column: 07
 Seq primer: M13RP1 reverse primer (ABI).
 Location/Qualifiers
 1..633
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="fgl0b07"
 /tissue_type="Iris"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /note="Organ: Eye; Vector: pCMVSPORT6; A human iris library (bx) was normalized by self-subtraction. One portion of double stranded plasmid DNA representing the library was linearized by NotI. This NotI digested library was used as a template for biotinylated RNA synthesis using SP6 RNA polymerase. Another portion of the double stranded plasmid library was converted to single-stranded circles in vitro using Gene II and Exonuclease III (Life Technologies). Single-stranded DNA (1 mg) was hybridized (Cot 500) with 41 mg of Bio-RNA and vector blocking oligonucleotides. The hybridized Bio-RNA/ss-circles were removed by streptavidin:phenol extraction. EST analysis was performed on the library at the NIH Intramural Sequencing Center (NISC)."

FEATURES
 source

ORIGIN

Alignment Scores:
 Pred. No.: 7,02e-101 Length: 633
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservativeness: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-657-740-1 (1-173) x CD672144 (1-633)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 52 ATGGATGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTCGGGCCCTTCTACCCACG 111
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 112 CGGCTGTTCGACCACTTTTTCGGCCAGGGCCCTTTTGTGATGATGACCTGCTGCCCTTCCTG 171
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 172 TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTTTCGGACCGTGTGACCTCCGGC 231
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValHisPhe 80
 Db 232 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTTCGTCATCTTCTCGATGTGAAGCACTTC 291
 Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 292 TCCCGGAGGACCTCACCCTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 351
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 Db 352 AACGAGCCGAGGACGACGACGCTACATTTTCGGTGTGATTCACCGCGCTACCGCTG 411
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140

|||||
 412 CGGTCCACGTGGACCGTGGCCCTCTCTGTCTCCCTGTCTGCGGATGCGTGCAC 471
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuaspAlaThrHisAlaGluArgAlaIlePro 160
 Db 472 TTCTGTGCCCCCAAGATCCAGACTGGCTGGATGTCACCCACGCGGAGCGCATCCCC 531

QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 532 GTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 570

RESULT 12

BM706270
 LOCUS 659 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-DW0-agh-k-09-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 UI-E-DW0-agh-k-09-0-UI 5', mRNA sequence.

BM706270
 VERSION

BM706270.1 GI:19019528

EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 659)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation by: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

FEATURES

source

1..659
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agh-k-09-0-UI"
 /tissue_type="lens"
 /dev_stages="adult"
 /lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT7T3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CGATTAGCGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.:

7.42e-101

Length:

659

Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM706270 (1-659)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 57 ATGGATGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTTACCCACG 116
 QY 21 ArgLeuPheAspGlnPhePheGlyGluLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 117 CGGCTGTTCGACCAAGTTTTTCGGCGAGGCGCTTTTGGATATGACCTGCTGCCCTTCCTG 176
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 177 TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTTTCGCGACCGTCTGGACTCCGCGC 236
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 237 ATCTCTCAGGTTCCATCCGACCGGACCAAGTTCGTCTATCTTCCTCGATGTGAAGCACTTC 296
 QY 81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
 Db 297 TCCCGGAGGACCTCACCGTGAAGGTGCAGACCACTTTGTGGAGATCCACGGAAGCAC 356
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
 Db 357 AACGAGCGCCAGGACGACCACTTATTCCTGAGTTCACCGCGCTGCTGCGCTG 416
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 417 CCGTCCAAACGTGGACAGTCCGCGCTCTCTTGTCTCTGCTGCGATGGCATGCTGACC 476
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 477 TTCTGTGCCCCCAAGATCCAGACTGGCTGGATGCGGCTGCGGAGCGGAGCATCCCC 536
 QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 537 GTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 575

RESULT 13

BM706270

LOCUS

DEFINITION

BM706270 Soares retina N2b4HR Homo sapiens cDNA clone

IMAGE:220285, mRNA sequence.

BM706270

VERSION

BM706270.1 GI:27882169

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

REFERENCE

1 (bases 1 to 681)

Ebert,L., Helli,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,

Radelof,U., Schneider,D. and Korn,B.

Human Unigeneset - RZPD3

Unpublished (2003)

JOURNAL

COMMENT

Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

RZPD; IMAGp998P14436.

RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

Human Unigeneset - RZPD3 (RZPDLIB No.972)

http://www.rzpd.de/CloneCards/cgi-

bin/showlib.pl.cgi/response?libNo=972 Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH

Heubnerweg 6, D-14059 Berlin, Germany

Tel: +49 30 32639 101

Fax: +49 30 32639 111

```

www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACACAGGAACAGCTATGAC.

FEATURES
    source
        Location/Qualifiers
            1..681
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IWAGp998P14436 ; IWAGR:220285"
                /sex="male"
                /tissue_type="retina"
                /dev_stage="55 year old"
                /lab_host="DH10B (ampicillin resistant)"
                /clone_lib="Soares retina N2b4HR"
                /note="Organ: eye; Vector: p7T73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTT 3'] ,
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T73 vector
(Pharmacia) . The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M.Fátima Bonaldo."
```

[illegible]

Qy	1	MetAspValThrIleGlnHisProTyrPheIysArgThrLeuGlyProPheTyrProSer	20
Db	93	ATGGAGCTGACCATCCAGCACCCCTGGTTCGAAGCGCACCCCTGGGGCCCTTCTACCCACG	152
Qy	21	ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu	40
Db	153	CGGCTGTTCACACAGTTTTTTCGGCGAGGGCTTTTTCAGTATGACCTGCTGCCCTTCTCTG	212
Qy	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60
Db	213	TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCGCGCACCGTGTGGAGCTCCGGC	272
Qy	61	IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe	80
Db	273	ATCTCTGAGGTTTCATCCGACCGGGACAGTTCGTCTATCTTCTCGATGTGAAGCACTTC	332
Qy	81	SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis	100
Db	333	TCCCCGGAGGACCTCACCGTGAAAGTGCAGGACGACCTTGTGGAGATCCACGGAAAGCAC	392
Qy	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
Db	393	AACGAGCGCCAGACGACGACCGGCTCATTTCCGTGAGTTTCACCCGCCCTTACCGCCTG	452
Qy	121	ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	140
Db	453	CCGTCCAAACGTGGACCATCGGCCCTCTCTTGTCTCCCTGTCTGCCGATGGCATGCTGACC	512
Qy	141	PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro	160
Db	513	TTCTGTGGCCCCAAGATCCAGATGGCTGGATGCCACCCACGCCGAGGAGGCATCCCC	572
Qy	161	ValSerArgGluGluLysProThrSerSerAlaProSerSer	173

Db	573 GTCTCGGGAGGAGAAGCCACCTCGGTCTCCTCGTCC	611
RESULT 14		
BF727324		
LOCUS	BF727324	698 bp mRNA linear EST 05-JAN-2001
DEFINITION	by91nh12.y1 Human lens cDNA (Un-normalized, unamplified): BY Homo sapiens cDNA clone by91nh12 5', mRNA sequence.	
ACCESSION	BF727324	
VERSION	BF727324.1	GI:12043235
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarhini; Homnidae; Homo.	
AUTHORS	1. (bases 1 to 698)	
TITLE	Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.	
JOURNAL	NIHBIANK: EST analysis and bioinformatics for ocular genomics Invest. Ophthalmol. Vis. Sci. 41 (2000) In press	
COMMENT	Contact: Wistow G Section on Molecular Structure and Function National Eye Institute 6/331, NIH, Bethesda, MD 20892-2740, USA Tel: 301 402 3452 Fax: 301 496 0078 Email: graeme@helix.nih.gov Plate: 19 row: h column: 12 Seq primer: M13RP1 reverse primer (ABI). Location/Qualifiers 1..698	
FEATURES		
source		

```

FEATURES
source
1. 639
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="by19h12"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
By"

```

Notes: Organ: Eye; Vector: pcwvSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pcwvSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript plasmid System full details of which are contained in the manufacturer's Instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGACGGCGCC('T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:			
Pred. No.:	8.03e-101	Length:	698
Score:	916.00	Matches:	173
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-10-657-740-1 (1-173) x BF727324 (1-698)

Qy	1	MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer	20
Db	55	ATGGAGCGTGACCATTCAGCACCCCTGGTTCAAGCGCACCTCGGGGCCCTTCTACCCGAC	114
Qy	21	ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu	40
Db	115	CGCGCTGTTGCACAGATTTTTCGCGAGGGGCCCTTTTGTAGTATGACCTGTCGCCCTTCCTG	174

QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 |||||
 Db 175 TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTCTTCGGACCGTGTGGACTCCGGC 234
 |||||
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 |||||
 Db 235 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTTCGTTCATCTTCTCGATGTGAAGCACTTC 294
 |||||
 QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 |||||
 Db 295 TCCCGGAGGACCTCACCCTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGAC 354
 |||||
 QY 101 AsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 |||||
 Db 355 AACGAGCGCCAGGACGACCGGCTACATTTCCGTGAGTTCACCGCGCTACCGCTG 414
 |||||
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 |||||
 Db 415 CGGTCCAAGTGCAGCAGTGGCCCTCTCTTGTCTCCCTGTCTGCGATGGCATGTGACC 474
 |||||
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 |||||
 Db 475 TTCTGTGCCCCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGAGCGACCATCCCC 534
 |||||
 QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 |||||
 Db 535 GTGTCCGGGAGGAGAGCCCACTTCGGCTCCCTCGTCC 573
 |||||

RESULT 15

BM722336
 LOCUS
 DEFINITION
 UI-E-E00-abx-i-06-0-UI r1 UI-E-E00 Homo sapiens cDNA clone

UI-E-E00-abx-i-06-0-UI 5', mRNA sequence.

BM722336

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

CNA Library preparation: Dr. M. Bento Soares, University of Iowa

CNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

CNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..724

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-E00-abx-i-06-0-UI"

/tissue_type="fetal eye"

/dev_stage="fetal"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-E00"

/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a

modified polylinker; Site_1: EcoR I; Site_2: Not I;

UI-E-E00 is a cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is CGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 8.44e-101 Length: 724
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM722336 (1-724)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
 |||||
 Db 51 ATGGACGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTGGGGCCCTTTACCCACG 110
 |||||
 QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 |||||
 Db 111 CGGCTGTTCCGACCACTTTTCGGCGAGGCGCTTTTGTAGTATGACCTGCTGCCCTTCCTG 170
 |||||
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 |||||
 Db 171 TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCCGACCGCTCTGGACTCCGGC 230
 |||||
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 |||||
 Db 231 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTCGTATCTTCTCGATGTGAAGCACTTC 290
 |||||
 QY 81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
 |||||
 Db 291 TCCCGGAGGACCTCACCCTGAAGTGCAGCACGACCTTTGTGGAGATCCACGGAAGCAC 350
 |||||
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 |||||
 Db 351 AACGAGCGCCAGGACGACCGCTACATTTCCGTGAGTTCACCGCGCTACCGCTG 410
 |||||
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 |||||
 Db 411 CCGTCCAACTGGACCACTCGGCCCTCTCTTGTCTGCTGCTGCGATGGCATGTGACC 470
 |||||
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 |||||
 Db 471 TTCTGTGCCCCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGAGCGACCATCCCC 530
 |||||
 QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 |||||
 Db 531 GTGTCCGGGAGGAGAAAGCCACCTCGGCTCCCTCGTCC 569
 |||||

RESULT 16

BM722779

LOCUS

DEFINITION

UI-E-E00-ahy-1-01-0-UI r1 UI-E-E00 Homo sapiens cDNA clone

UI-E-E00-ahy-1-01-0-UI 5', mRNA sequence.

BM722779

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homosapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

FEATURES

source

559 bp mRNA linear EST 01-MAR-2002

REFERENCE 1 (bases 1 to 559)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
 1..559
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-E00-ahy-1-01-0-UI"
 /tissue_type="fetal eye"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-E00"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-E00 is a cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:
 Pred. No.: 1.82e-100 Length: 559
 Score: 912.00 Matches: 172
 Percent Similarity: 99.42% Conservative: 0
 Best Local Similarity: 99.42% Mismatches: 1
 Query Match: 99.56% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM722779 (1-559)

Qy 1 MetAspValThrIleGlnHisProThrPhelysArgThrLeuGlyProPheTyProSer 20
 Db 36 ATGGATGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCAGC 95
 Qy 21 ArgLeuPheAspGlnPheGlyGlyLeuPheGluTyArgPheLeuProPheLeu 40
 Db 96 CGGCTGTTGACCAAGTTTTCGGCGAGGGCCCTTTTGTAGTATGACCTGCTGCCCTTCTG 155
 Qy 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 156 TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTCTTCGCGACCGTGTGACTCCGGC 215
 Qy 61 IleSerGluValArgSerAspArgAspIlePheLeuAspValIleHisPhe 80
 Db 216 ATCTCTGAGTTTCGATCCGACGGGACAAGTTTCGTTCATCTTCTCGATGTGAAGCACTTC 275

Qy 81 SerProGluAspLeuThrValIysValGlnAspPheValGluIleHisGlyIysHis 100
 Db 276 TCCCGGAGGACCTCAGCGTGAAGGTGAGGACGACTTTGTGGAGATCCACGAAAGCAC 335
 Qy 101 AsnGluArgGlnAspAspHisGlyTyTyrIleSerArgGluPheHisArgGlyrArgLeu 120
 Db 336 AACGAGCCGAGGAGGACGACCGGTACATTTCCTGAGTTCCACCGCGCTACCGCTG 395
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 396 CGTCCACGTTGACCACTCGGCCCTCTTGTCTCCCTGTCTGCCGATGGCATGCTGACC 455
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 456 TTCTGTGCGCCCAAGATCCAGACTGGCTGGATGCCACCCACGCGAGGAGCCATCCCC 515
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 516 GTGTGCGGGAGGAGAGCCACCTCGGNTCCCTCGTCC 554

RESULT 17

BM697066
 LOCUS BM697066 635 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-DW0-agm-g-03-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 ACCESSION UI-E-DW0-agm-g-03-0-UI 5', mRNA sequence.
 VERSION BM697066
 KEYWORDS EST.
 SOURCE BM697066.1 GI:19010324
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 635)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..635
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agm-g-03-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The

FEATURES

Location/Qualifiers
 1..635
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agm-g-03-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The

oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CGATTAGCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 2,878-100 Length: 635
 Score: 911.00 Matches: 172
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.45% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM697066 (1-635)

QY 2 AspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSerArg 21
 DB 1 GATGTGACCATTCAGACCCCTGGTTCAGCGCACCTGGGGCCCTTCTACCCAGCCGG 60
 QY 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSer 41
 DB 61 CTGTTTCGACCACTTTTCGGCGAGGGCCCTTTTGTAGTATGACCTGCTGCTTCTCTGTCG 120
 QY 42 SerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIle 61
 DB 121 TCCACCATCAGCCCTACTACCGCCAGTCCCTTTCGCGACCGTCTGGACTCGCGCATC 180
 QY 62 SerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSer 81
 DB 181 TCTGAGGTTGATCCGACCGGACAGTTCGTCACTTCTCGATGTGAGCATCTTCTCC 240
 QY 82 ProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsn 101
 DB 241 CCGGAGGACCTCACCCTGAAGGTGAGGACGACTTGTGGAGATCCACGGAAGCACAAC 300
 QY 102 GluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuPro 121
 DB 301 GAGCGCCAGGACGACCGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCGG 360
 QY 122 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 141
 DB 361 TCCACGTGGACCACTCGGCCCTCTCTGCTCCCTGCTGCGATGGCATGCTGACCTTC 420
 QY 142 CysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 161
 DB 421 TGTGCCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGAGCGAGCATGCCCGCTG 480
 QY 162 SerArgGluGluLysProThrSerAlaProSerSer 173
 DB 481 TCGCGGAGGAGAGAGCCACCTCGGCTCCCTCGTCC 516

RESULT 18

BF726358

LOCUS

DEFINITION BF726358 532 bp mRNA linear EST 05-JAN-2001
 sapiens cDNA clone by05d12 5', mRNA sequence.

ACCESSION

BF726358

VERSION

BF726358.1

KEYWORDS

EST

SOURCE

Homo sapiens

ORGANISM

REFERENCE

1 (bases 1 to 532)

AUTHORS

Wistow G.J., Bernstein S., Behal A. and Smith D.

TITLE

NEIBANK: EST analysis and bioinformatics for ocular genomics

JOURNAL

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

COMMENT

Section on Molecular Structure and Function

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 05 row: d column: 12

Seq primer: M13RP1 reverse primer (ABI).

FEATURES

Location/Qualifiers

1..532
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="by05d12"
 /tissue_type="Lens"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human Lens cDNA (Un-normalized, unamplified): BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTGTCTAGATCGAGCGCCCTTT15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 2,988-100 Length: 532
 Score: 910.00 Matches: 172
 Percent Similarity: 99.42% Conservatives: 0
 Best Local Similarity: 99.42% Mismatches: 1
 Query Match: 99.34% Indels: 0
 DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF726358 (1-532)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
 DB 2 ATGATGTGACCATTCAGACCCCTGGTTCAGCGCACCTGGGGCCCTTCTACCCAGC 61
 QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 DB 62 CGGCTGTTTCGACCACTTTTCGGCGAGGGCCCTTTTGTAGTATGACCTGCTGCTTCTG 121
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 DB 122 TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTTTCGCGACCGTCTCGACTCCGGC 181
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 DB 182 ATCTCTGAGGTTTCATTCGACCGGACAGTTCGTCTCATCTTCTTCGATGTGAGCAGCTTC 241
 QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 DB 242 TCCCGGAGGAGCTCACCCTGAAGGTGCAGGACGACATTTGTGGAGATCCACGGAAGCAC 301
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 DB 302 AACGAGCGCCAGGACGACACCGCTACATTTCCCGTGAAGTTCACCGCGCTACCGCGCTG 361
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 DB 362 CCGTCCAACTGGACCAAGTTGGCCCTCTCTTGTCTCCCTGCTGCGATGGCATGCTGACC 421
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160

```

Db      422 TTCTGTGGCCCCAAGATCCAGACTGGCCTGGATGTCACCCAGCGGAGGACCATCCCC 481
Qy      161 ValSerArgGluGluProThrSerAlaProSerSer 173
Db      482 GTGTGCGGAGGAGAGCCACCTCGGCTCCTCGTCC 520

RESULT 19
AY419530
LOCUS   Pan troglodytes CRYAA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY419530
VERSION   AY419530.1 GI:39775487
KEYWORDS  Pan troglodytes (chimpanzee)
SOURCE    Pan troglodytes
ORGANISM  Pan troglodytes (chimpanzee)
REFERENCE 1 (bases 1 to 522)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
          Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
          Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
          Adams,M.D. and Cargill,M.
          Inferring nonneutral evolution from human-chimp-mouse orthologous
          gene trios
JOURNAL   Science 302 (5652), 1960-1963 (2003)
PUBMED    14671302
REFERENCE 2 (bases 1 to 522)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
          Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
          Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
          Adams,M.D. and Cargill,M.
          Direct Submission
          Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
          Rockville, MD 20850, USA
TITLE     These sequences were made by sequencing genomic exons and ordering
          them based on alignment.
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT   These sequences were made by sequencing genomic exons and ordering
          them based on alignment.
FEATURES
source
1..522
Location/Qualifiers
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>522
/gene="CRYAA"
/locus_tag="HGM6921"
gene
Alignment Scores:
Pred. No.: 3.84e-100 Length: 522
Score: 909.00 Matches: 172
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 1
Query Match: 99.24% Indels: 0
Db: 9 Gaps: 0

US-10-657-740-1 (1-173) x AY419530 (1-522)

Qy      1 MetAspValThrIleGlnHisProThrPheLysArgThrLeuGlyProPheTyrProSer 20
Db      1 ATGACGTGACCATTCACGACCCCTGGTTCAGCGCACCTCGGGGCCCTTCTACCCACG 60
Qy      21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db      61 CGGTGTTTCGACCACTTTTTCGGGAGGAGGCTTTTGTAGTATGACCTCTGCTCTCTCTG 120
Qy      41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db      121 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCGACCGCTCTGCTGCTCGGC 180
Qy      61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPhe 80
Db      181 ATCTCTGAGGTTCCATCCGCGGACAAAGTTTCGTCACTCTTCTCGATGTGAAGCACTTC 240

```

```

Qy      81 SerProGluAspLeuThrValIlyValGlnAspAspPheValGluIleHisGlyLysHis 100
Db      241 TCCCGGAGGACCTCACCGTGAAGGTGCAGGACGACTTTGTGGAGATCCACGAAAGCAC 300
Qy      101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db      301 AACGAGCGCCAGNNNGACCGCTACATTTCCCGTGAGTTCCACCGCGGCTACCGCCTG 360
Qy      121 ProSerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db      361 CCGTCCACAGGTGGACAGTCGCGCCTCTCTTGTCTCCCTGTCGCCGATGGCATGCTGACC 420
Qy      141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db      421 TTCTGTGGCCCAAGATCCAGACTGGCTGATGCCACCCAGCGGAGGACCATCCCC 480
Qy      161 ValSerArgGluGluProThrSerAlaProSerSer 173
Db      481 GTGTGCGGAGGAGAGCCACCTCGGCTCCTCTCGTCC 519

RESULT 20
BF726253
LOCUS   sapiens cDNA clone by03h05 5', mRNA sequence.
DEFINITION sapiens cDNA clone by03h05 5', mRNA sequence.
ACCESSION BF726253
VERSION   BF726253.1 GI:12042164
KEYWORDS  sapiens (human)
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 577)
AUTHORS   Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
TITLE     NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL   Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT   Contact: Wistow G
          Section on Molecular Structure and Function
          National Eye Institute
          6/331, NIH, Bethesda, MD 20892-2740, USA
          Tel: 301 402 3452
          Fax: 301 496 0078
          Email: graeme@helix.nih.gov
          Plate: 03 row: h column: 05
          Seq primer: M13Rp1 reverse primer (ABI).
          Location/Qualifiers
FEATURES
1..577
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="by03h05"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDHI08"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
By"
/notes="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ng of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life
Technologies, essentially following the protocols of the
SuperScript Plasmid System full details of which are
contained in the manufacturer's Instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTTACGACGCGCGCCCTT15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."
```

Alignment Scores:

```

Pred. No.: 7,72e-100 Length: 577
Score: 907.00 Matches: 172
Percent Similarity: 99.42% Conservatives: 0
Best Local Similarity: 99.42% Mismatches: 1
Query Match: 99.02% Indels: 0
DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF726253 (1-577)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
Db 48 ATGGAGCGTACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCAGC 107
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyArgLeuLeuProPheLeu 40
Db 108 CGGCTGTTTCGACAGTTTTCGGCAGGCGCTTTTGTAGATGAGCTCTGCGCTTCCTG 167
Qy 41 SerSerThrIleSerProTyTrpArgGlnSerLeuPhePheArgThrValLeuAspSerGly 60
Db 168 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTTTCGACCGCTGCTGGACTCCGGC 227
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db 228 ATCTCTGAGGTTTCGATCCGACCGGACCAAGTTTCGTCTCTTCGTGATGAGCACTTC 287
Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
Db 288 TCCCGGAGGACCTCAGCGTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 347
Qy 101 AsnGluArgGlnAspAspHisGlyTyTrpIleSerArgGluPheHisArgArgTyArgLeu 120
Db 348 AACGAGCGCCAGGACGACCGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTG 407
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 408 CGGTCCAAAGTGGACAGTCCGCGCCCTCTCTGTCTCCCTGCTGCGCGATGGCATGTGACC 467
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 468 TTCTGTGGCCCCAAGATCCAGACTGGCCTGGATGCCACCGCGCGAGGAGCCATCNC 527
Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 528 GTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 566

```

```

RESULT 21
BF726422 587 bp mRNA linear EST 05-JAN-2001
LOCUS by06d05.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
DEFINITION sapiens cDNA clone by06d05 5', mRNA sequence.
ACCESSION BF726422
VERSION BF726422.1 GI:12042333
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 587)
Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
NEIBANK: EST analysis and bioinformatics for ocular genomics
Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 06 row: d column: 05
Seq primer: M13RF1 reverse primer (ABI).
Location/Qualifiers
1..587
/organism="Homo sapiens"
FEATURES
source

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="by06d05"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
BY"
/notes="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life
Technologies, essentially following the protocols of the
SuperScript Plasmid System full details of which are
contained in the manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTCTAGATCGGAGCGCGCC(T)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 7,91e-100 Length: 587
Score: 907.00 Matches: 172
Percent Similarity: 99.42% Conservatives: 0
Best Local Similarity: 99.42% Mismatches: 1
Query Match: 99.02% Indels: 0
DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF726422 (1-587)

```

```

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
Db 8 ATGGAGCGTACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCAGC 67
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyArgLeuLeuProPheLeu 40
Db 68 CGGCTGTTTCGACAGTTTTCGGCAGGCGCTTTTGTAGATGAGCTCTGCGCTTCCTG 127
Qy 41 SerSerThrIleSerProTyTrpArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 128 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTTTCGACCGTGTGAGCTCCGSC 187
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db 188 ATCTCTGAGGTTTCGATCCGACCGGACCAAGTTCGTCTCTCTCGATGTGAAGCACTTC 247
Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
Db 248 TCCCGGAGGAGCTCACCCTGAGGTGAGGAGCACTTTGTGGAGATCCCGAAGCAC 307
Qy 101 AsnGluArgGlnAspAspHisGlyTyTrpIleSerArgGluPheHisArgTyArgLeu 120
Db 308 AACGAGCGCCAGGACGACCGCTACATTTCCCTGAGTTCACCGCGCTACCGCTG 367
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 368 CCGTCCAAAGTGGACCAAGTTCGCGCCCTCTCTTGTCTCCCTGTGCGATGGCATGTGACC 427
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 428 TTCTGTGGCCCCAAGATCCAGACTGGCCTGGATGCCACCGCGAGGAGCCATCNC 487
Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 488 GTGTGCGGGAGGAGAGACCCACCTCGGCTCCCTCGTCC 526

```

```

RESULT 22
BF726330
LOCUS BF726330

```

```

597 bp mRNA linear EST 05-JAN-2001

```

by05b01.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
 sapiens cDNA clone by05b01 5', mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 EST.
 BF726330.1 GI:12042241
 Homo sapiens (human)
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS
 Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
 TITLE
 NEIBANK: EST analysis and bioinformatics for ocular genomics
 JOURNAL
 Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
 COMMENT
 Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 05 row: b column: 01
 Seg primer: M13Rpl reverse primer (ABT).

```

Location/Qualifiers
1. .597
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="by05b01"
/tissue_types="Lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
By"
/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life
Technologies, essentially following the protocols of the
SuperScript Plasmid System full details of which are
contained in the manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTCTCATGCGAGCGCGCCC(T)15-3']. Not I/blunt
end inserts were cloned into the Not I/ScoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

```

ORIGIN

Alignment Scores:					
Pred. No.:	1.07e-99	Length:	597		
Score:	906.00	Matches:	171		
Percent Similarity:	98.84%	Conservative:	0		
Best Local Similarity:	98.84%	Mismatches:	0		
Query Match:	98.91%	Indels:	0		
DB:	2	Gaps:	0		
US-10-657-740-1 (1-173) x BF726330 (1-597)					
Qy	1	MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer	20		
Dd	2	ATCGATGTGCACCATCAGCACCCCTGGTTCAAGCGCACCTTGSGGCCCTCTACCCCAGC	61		
Qy	21	ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuPropheLeu	40		
Dd	62	CGGCTGTTCACCAGATTTCGSCGAGGGCTTTTCAGTAGTACCTGCTGCCTTCCTG	121		
Qy	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60		
Dd	122	TGCTGCACCATCAGCCCCCTACTACCGCCAGTCCCTCTTCGCGACCGTGTGGACTCCGGC	181		
Qy	61	IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe	80		

Db	182	ATCTCTGAGGTTTCGATCCGAGCGGACAAGTTTCGTTCATCTTCTTCGATGTGAAGCACTTC	241
Qy	81	Ser:ProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis	100
Db	242	TCCCGGAGGACTTCACCGTGAAGGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC	301
Qy	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
Db	302	AACGAGCGCCAGACGACGACGCTACTTTCCTCGTGAAGTTCACCGCGCTACCGCGCTG	361
Qy	121	ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	140
Db	362	CCGTCCAACTGACGACGTCGGCCCTCTCTTGTCTCCCTGTCTCCGATGCACTGTCATC	421
Qy	141	PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro	160
Db	422	TTCTGTGGCCCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGAGCGGCATCCCC	481
Qy	161	ValSerArgGluGluLysProThrSerAlaProSerSer	173
Db	482	GTCTCGCGGAGGAGAAGCCACCTCGGCTCCCTCGTCC	520
RESULT 23			
LOCUS	BM706139	580 bp	mRNA linear EST 28-FEB-2002
DEFINITION	UI-E-DW0-agg-b-16-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone		
SOURCE	UI-E-DW0-agg-b-16-0-UI 5', mRNA sequence.		
ACCESSION	BM706139.1		
VERSION	GI:19019397		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 580)		
TITLE	Bonaldio,M.F., Lennon,G. and Soares,M.B.		
JOURNAL	Normalization and subtraction: two approaches to facilitate gene		
MEDLINE	discovery		
PUBMED	Genome Res. 6 (9), 791-806 (1996)		
COMMENT	97044477 889548 Contact: Soares, MB Coordinated Laboratory for Computational Genomics University of Iowa 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: bento-soares@uiowa.edu Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Genetics (www.resgen.com). Seq primer: M13 Reverse. Location/Qualifiers 1. 580 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="UI-E-DW0-agg-b-16-0-UI" /tissue_type="lens" /dev_stage="adult" /lab_host="DH10B (Life Technologies) (T1 phage resistant)" /clone_lib="UI-E-DW0" /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldio, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not		
FEATURES			
source			

I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:
Pred. No.: 2,39e-99 Length: 580
Score: 903.00 Matches: 170
Percent Similarity: 99.42% Conservatives: 0
Best Local Similarity: 99.42% Mismatches: 1
Query Match: 98.58% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM706139 (1-580)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
Db ATGGATGTGACCATCCAGACCCCTGGTTCAGCGCACCCCTGGGGCCCTTCTACCCCGAC 115
Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db CGGCTGTTTCGACCACTTTTCGGCGAGGGCCCTTTTGATGATGACTGCTGCCCTTCCTG 175
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTTTCGACACCGTCTGGACTCCGGC 235
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db ATCTCTGAGGTTCGATCCGACCGGACCAAGTTCGTCATCTTCTCGATGTGAAGCACTTC 295
Qy 81 SerProGluAspLeuThrValLysValGlnAspPhePheValGluIleHisGlyLysHis 100
Db TCCCGGAGGACCTCACCGTGAAGTGCAGGACGACTTGTGGAGATCCACGGAAAGCAC 355
Qy 101 AsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db AACGAGCGCCAGGACGACCGCTTACATTTCCCGTGATTCACACCGCGCTACCGCTG 415
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db CGGTCCACGTGACGACGTCGGCCCTCTCTGCTCNCCTGCTGCGGATGGATGCTGACC 475
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db TTCTGTGCCCCAAGATCCAGACTGGGCTGGATGCCACCCAGCGCGAGGAGCATCCCC 535
Qy 161 ValSerArgGluGluLysProThrSerAlaPro 171
Db GTGTCCGCGAGGAGAGCCCAACTCGGTCC 568

RESULT 24
BM722650
LOCUS
DEFINITION
UI-E-E00-ahy-a-10-0-UI.r1 UI-E-E00 Homo sapiens cDNA clone
UI-E-E00-ahy-a-10-0-UI 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 586)
REFERENCE
AUTHORS
TITLE
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477

PUBMED
COMMENT

8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reagen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
1. 586
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-E00-ahy-a-10-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-E00"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-E00 is a cDNA library containing the following
tissue(s): fetal eye. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGGTATACC. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

FEATURES
source

ORIGIN

Alignment Scores:
Pred. No.: 2,42e-99 Length: 586
Score: 903.00 Matches: 173
Percent Similarity: 99.43% Conservatives: 0
Best Local Similarity: 99.43% Mismatches: 0
Query Match: 98.58% Indels: 1
DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM722650 (1-586)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
Db ATGGATGTGACCATCCAGACCCCTGGTTCAGCGCACCCCTGGGGCCCTTCTACCCCGAC 109
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db CGGCTGTTTCGACCACTTTTCGGCGAGGGCCCTTTTGATGATGACTGCTGCCCTTCCTG 169
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTTTCGACACCGTCTGGACTCCGGC 229
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db ATCTCTGAGGTTCGATCCGACCGGACCAAGTTCGTCATCTTCTCGATGTGAAGCACTTC 289
Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
Db TCCCGGAGGACCTCACCGTGAAGTGCAGGACGACTTTTGTGGAGATCCACGGAAAGCAC 349
Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120

```

Db      350 AACGAGCGCCAGACGACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTG 409
Qy      121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db      410 CGGTCCACAGTGGACCAAGTCGGCCCTCTTGCTCCCTGCTCGCGATGGATGCTGACC 469
Qy      141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThr-HisAlaGluArgAlaIlePr 160
Db      470 TTCTGTGGCCCAAGATCCAGACTGGCTGGATGCCACCCACGCGGAGCGGACCATCCC 529

Qy      160 oValSerArgGluGluLysProThrSerAlaProSerSer 173
Db      530 CGTGTGCGGAGGAGGAGACCCACCTCGCTCCCTCGTCC 569

RESULT 25
LOCUS   BM697368
DEFINITION UI-E-DW0-agm-d-06-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
VERSION   BM697368
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 569)
          Bonaldo,M.F., Lennon,G. and Soares,M.B.
          Normalization and subtraction: two approaches to facilitate gene
          discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
          Coordinated Laboratory for Computational Genomics
          University of Iowa
          375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
          Tel: 319 335 8250
          Fax: 319 335 9565
          Email: bento-soares@uiowa.edu
          Tissue Procurement: Dr. Gregg Hageman
          cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Researchers may obtain clones from Research
          Genetics (www.reggen.com).
          Seq primer: M13 Reverse.
          Location/Qualifiers
            1..569
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="UI-E-DW0-agm-d-06-0-UI"
              /tissue_type="lens"
              /dev_stage="adult"
              /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
              /clone_lib="UI-E-DW0"
              /note="Organ: eye; Vector: pTT73-Pac (Pharmacia) with a
              modified polylinker; Site_1: EcoR I; Site_2: Not I;
              UI-E-DW0 is a cDNA library containing the following
              tissue(s): lens. The library was constructed according to
              Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
              1996. First strand cDNA synthesis was primed with an
              oligo-dT primer containing a Not I site. Double stranded
              cDNA was ligated to an EcoR I adaptor, digested with Not.
              I, and cloned directionally into pTT73-Pac vector. The
              oligonucleotide used to prime the synthesis of
              first-strand cDNA contains a library tag sequence that is
              located between the Not I site and the (dT)18 tail. The
              sequence tag for this library is CGATTAGCGA. This library
              was created for the program, Gene Discovery in the Visual
              System, supported by National Eye Institute (NEI)."
```

```

ORIGIN
Alignment Scores:
Pred. No.: 4.08e-99 Length: 569
Score: 901.00 Matches: 170
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.36% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM697368 (1-569)

Qy      4 ThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPhe 23
Db      2 ACCATCCAGCACCCCTGGTTTCAAGCGCACCCCTGGGGCCCTTCTATCCCGCGGGTGTTC 61
Qy      24 AspGlnPhePheGlyGluGlyLeuPheGluTyrArgLeuLeuProPheLeuSerSerThr 43
Db      62 GACCAAGTTTTTCGGCGAGGGCCCTTTTGGATATGACCTGCTGCCCTTCTGTGTCACCC 121
Qy      44 IleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGlu 63
Db      122 ATCAGCCCTACTACCGCCAGTCCCTCTTCGCGACCGTCTGGACTCCGCGCATCTCTGAG 181
Qy      64 ValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGlu 83
Db      182 GTTCGATCCGACCGCGGACAGTTCGTCATCTTCCTCGATGTGAAGCACTTCTCCCGGAG 241
Qy      84 AspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGluArg 103
Db      242 GACCTCACCGTGAAGTGCAGGACACTTTGTGAGATCCACGGAAGCAACACGAGCGC 301
Qy      104 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 123
Db      302 CAGGACGACCAAGCTACATTTCCCGTGAGTTCACCGCGCTACCGCTCCGCTCCAC 361
Qy      124 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 143
Db      362 GTGACCAAGTGGGCCCTCTCTTGTCTGCTGTCGCGATGCGATGCTGACCTTCTGTGGC 421
Qy      144 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 163
Db      422 CCCAAGATCCAGACTGCGCTGGATGCCACCCACCGCGAGCGGATCCCGTGTGCGGG 481
Qy      164 GluGluLysProThrSerAlaProSerSer 173
Db      482 GAGGAGAGAGCCCACTCGGCTCCCTCGTCC 511

RESULT 26
LOCUS   BM697160
DEFINITION UI-E-DW0-agm-i-06-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
VERSION   BM697160
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     1 (bases 1 to 661)
          Bonaldo,M.F., Lennon,G. and Soares,M.B.
          Normalization and subtraction: two approaches to facilitate gene
          discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
          Coordinated Laboratory for Computational Genomics
          University of Iowa
          375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
          Tel: 319 335 8250
          Fax: 319 335 9565
```

Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
 1..661

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-aggm-i-06-0-UI"
 /tissue_type="lens"
 /dev_stages="adult"
 /lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DW0 is a cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 5,01e-99 Length: 661
 Score: 901.00 Matches: 170
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 98.36% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM697160 (1-661)

Qy 4 ThrileGlnHisProTyrPhelysArgThrLeuGlyProPheTyrProSerArgLeuphe 23
 Db 3 ACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTTCTACCCCGCGGTGTC 62
 Qy 24 AspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThr 43
 Db 63 GACCAAGTTTTTCGGCGAGGGCTTTTTCAGTATGACCTGCTGCCCTTCTGTCTCCACC 122
 Qy 44 IleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGlu 63
 Db 123 ATCAGCCCTCTACTACCGCCAGTCCCTCTTCGCGACCGTGTGGACTCCGGCATCTCTGAG 182
 Qy 64 ValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGlu 83
 Db 183 GTTCATCCGACCGCGACAAGTTCTCATCTTCTTCGATGTGAACACTTCTCCCGCGAG 242
 Qy 84 AspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArg 103
 Db 243 GACCTCACCGTGAAGTGCAGGACGACATTTGTGGAGATCCACGGNAACACACGAGGCG 302
 Qy 104 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsn 123
 Db 303 CAGGACGACACCGGCTACATTTCCCGTTCAGTTCCACCGCGCTACCGCTCGCGTCCAAAC 362
 Qy 124 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 143
 Db 363 GTGGACCACTCGGCCCTCTCTTGTCTCCCTGTCTGCGGATGGCATGCTGACCTTCTGTGGC 422

Qy 144 ProlysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 163
 Db 423 CCCAAGATCCAGATGCGCTGGATGCCACCCACCGCGAGCGGCATCCCCGTGTCTGGCG 482
 Qy 164 GluGluLysProThrSerAlaProSerSer 173
 Db 483 GAGGAGAAGCCACCTCGCTCCCTCGTCC 512

RESULT 27

BO640267

LOCUS

DEFINITION

BO640267

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

SECTION

NATIONAL

6/331, NIH

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1..604

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="he26c10"

/tissue_type="Retina"

/dev_stage="Adult"

/lab_host="EMDH10B"

/clone_lib="Human Retina cDNA (Un-normalized, unamplified): hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library in the construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor

[5'-pGATGTTCTGATCGCGCGCGCCC(T)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 5,86e-99 Length: 604

Score: 900.00 Matches: 170

Percent Similarity: 98.27% Conservative: 0

Best Local Similarity: 98.27% Mismatches: 3

Query Match: 98.25% Indels: 0

DB: 5 Gaps: 0

source

BO640267

LOCUS

DEFINITION

BO640267

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

SECTION

NATIONAL

6/331, NIH

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1..604

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="he26c10"

/tissue_type="Retina"

/dev_stage="Adult"

/lab_host="EMDH10B"

/clone_lib="Human Retina cDNA (Un-normalized, unamplified): hd/he"

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library in the construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor

[5'-pGATGTTCTGATCGCGCGCGCCC(T)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 5,86e-99 Length: 604

Score: 900.00 Matches: 170

Percent Similarity: 98.27% Conservative: 0

Best Local Similarity: 98.27% Mismatches: 3

Query Match: 98.25% Indels: 0

DB: 5 Gaps: 0

US-10-657-740-1 (1-173) x BQ640267 (1-604)

```

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
DB 47 ATGGATGTGACCATCAGACACCCCTGGTGAAGCGCACCTGGGGCCCTTCTACCCAGC 106
QY 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGlyTyAspLeuLeuProPheLeu 40
DB 107 CGCGTGTTCACACAGTTTTCGGCGAGGGCCCTTTTGGATGATGACCTGCTGCCCTTCCTG 166
QY 41 SerSerThrIleSerProTyTrpArgGlnSerLeuPheArgThrValLeuAspSerGly 60
DB 167 TCGTCCACCATCAGCCCTACTACCGCAGTCCCTCTTCGGACCGTGTGACTCCGCT 226
QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
DB 227 TTTCTTTGAGGTTTCGATCCGACCGGACAAAGTTTCGTCTCTCGATGGAAGCACTTC 286
QY 81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
DB 287 TCCCGGAGGACCTCACCCTGAAGGTGACGAGCACTTTGTGGAGATCCACGGAAGCAC 346
QY 101 AsnGluArgGlnAspAspHisGlyTyTrpIleSerArgGluPheHisArgArgTyArgLeu 120
DB 347 AACGAGCGCCAGACACCGCTACATTTCCGAGATTCACCGCCGCTACCGCCCTG 406
QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
DB 407 CGGTCCAAAGTGGACAGTCGGCCCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 466
QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
DB 467 TTTCTGTGGCCCAAGATCCAGACTGGCCCTGGATGCCACCCAGCGAGCGAGCCATCC 526
QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
DB 527 GTGTCCGGGAGAGAACCCACCTTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 565

```

RESULT 28

BF727002 577 bp mRNA linear EST 05-JAN-2001
 LOCUS by15c11.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
 DEFINITION sapiens cDNA clone by15c11 5', mRNA sequence.

ACCESSION BF727002
 VERSION BF727002.1 GI:12042913
 KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 577)

Wistow G.J., Bernstein S., Behal A. and Smith, D.
 NEIBANK: EST analysis and bioinformatics for ocular genomics
 Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/31, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 15 row: c column: 11

Seq primer: M13Rpl reverse primer (ABI).

Location/Qualifiers

1..577

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="by15c11"

/tissue_type="Lens"

/dev_stage="Adult"

/lab_host="EMDHI08"

/clone_lib="Human Lens cDNA (Un-normalized, unamplified):"

BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the Superscript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGATAGTTCTAGATCCGAGCGGCCCT(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores: 9.65e-99 Length: 577

Pred. No.: 898.00 Matches: 170

Score: 99.42% Conservative: 0

Best Local Similarity: 99.42% Mismatches: 1

Query Match: 98.03% Indels: 0

DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF727002 (1-577)

```

QY 3 ValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSerArgLeu 22
DB 1 GTGACATCCAGCACCCCTGGTTCAAGCGCACCTGNGGCCCTTCTACCCAGCGGCTG 60
QY 23 PheAspGlnPhePheGlyGluGlyLeuPheGluTyArgLeuLeuProPheLeuSerSer 42
DB 61 TTCGACCAAGTTTTTCGGCGAGGGCTTTTGTAGTATGACCTGCTGCTCTCTCTCTCT 120
QY 43 ThrIleSerProTyTrpArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSer 62
DB 121 ACCATCAGCCCTTACTACCGCCAGTCCCTTTCGACCGCTGCTGACCTCCGCACTCT 180
QY 63 GluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerPro 82
DB 181 GAGGTTCGATCCGACCGGAGACAGTTCGTCTCTCTCGATGTGAGAGCACTTCTCCCG 240
QY 83 GluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGlu 102
DB 241 GAGGACCTCACCGTGAAGGTGCAGGACGACTTTGTGGAGATCCACGGAAGCACACGAG 300
QY 103 ArgGlnAspAspHisGlyTyTrpIleSerArgGluPheHisArgArgTyArgLeuProSer 122
DB 301 CGCCAGGACGACACCGCTACATTTCCCGTGAGTTTCCACCGCTGCTGCGCTGCGCTCC 360
QY 123 AsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCys 142
DB 361 AACGTGGACCAAGTCCGACCTCTCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 143 GlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSer 162
DB 421 GSCCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGGAGGAGCCATCCCGGTGTCG 480
QY 163 ArgGluGluLysProThrSerAlaProSerSer 173
DB 481 CGGAGGAGAGAGCCACCTCCGCTCCCTCGCTCCCTCGCTCCCTCGCTCCCTCGCTCC 513

```

RESULT 29

BF726854

LOCUS

DEFINITION

by13a10.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo

ACCESSION BF726854

VERSION BF726854.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

BF726854 518 bp mRNA linear EST 05-JAN-2001

by13a10.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo

sapiens cDNA clone by13a10 5', mRNA sequence.

GI:12042765

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 518)
AUTHORS Wisnow.G.J., Bernstein.S., Behal.A. and Smith.D.
TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
COMMENT Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 13 row: a column: 10

Seq primer: M13RPI reverse primer (ABI).

FEATURES
source

1..518
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="by13a10"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH108"
/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life
Technologies, essentially following the protocols of the
SuperScript Plasmid System full details of which are
contained in the manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTCCTAGATCGGAGCGGCC(T)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
Pred. No.: 1 46e-98 Length: 518
Score: 896.00 Matches: 169
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.82% Indels: 0
DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF726854 (1-518)

QY 5 IleGlnHisProTrpPhelysArgThrLeuGlyProPheTyProSerArgLeuPheAep 24
Db 2 ATCCAGCACCCCTGGTTCAAGCGCACCTCTGGGGCCCTTCTACCCAGCGCGTGTTCGAC 61
QY 25 GlnPhePheGlyGluGlyLeuPheGluTyAspLeuLeuProPheLeuSerSerThrIle 44
Db 62 CAGTTTTTCGGGAGGGCCCTTTTGTAGTATGACCTGCTGCCCTTCCTGTCGTCACCAATC 121
QY 45 SerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluVal 64
Db 122 AGCCCTTACTACCGCCAGTCCCTCTTCCGACCGTGTGAGCTTCGGCATCTCTGAGGTT 181
QY 65 ArgSerAspArgAspIlyPheValIlePheLeuAspValIlyHisPheSerProGluAsp 84
Db 182 CGATCCGACGGGACCAAGTTGGTTCATCTTTCGATGTGAAGCACTTCTCCCGGAGGAC 241
QY 85 LeuThrValIlyValGlnAspAspPheValGluIleHisGlyIlyHisGlnArgGln 104
Db 242 CTCACCGTGAAGTGCAGACGACGACTTTGTGGAGATCCACGGAAGCACACGAGCGCCAG 301
QY 105 AspAspHisGlyTyIleSerArgGluPheHisArgTyArgLeuProSerAsnVal 124

Db 302 GACGACCACGGCTACATTTCCGTCGAGTTCACCGCGCTACCGCTCGCTCCACCGTG 361
QY 125 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 144
Db 362 GACCAGTCGGCCCTCTCTTGTCTCCCTGCTGCGGATGGCATGCTGACCTTCTGTGGCCCC 421
QY 145 LysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGlu 164
Db 422 AAGATCCAGACTGGCTGATGCCACCCAGCGGAGCGGAGCATCCCGTGTGCGGGAG 481
QY 165 GluLysProThrSerAlaProSerSer 173
Db 482 GAGAAGCCACCTCGGCTCCCTCGTCC 508

RESULT 30

BF726438

LOCUS

DEFINITION

by06f01.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo

523 bp mRNA linear EST 05-JAN-2001

sapiens cDNA clone by06f01 5', mRNA sequence.

ACCESSION

BF726438

VERSION

BF726438.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 523)

AUTHORS

Wistow.G.J., Bernstein.S., Behal.A. and Smith.D.

TITLE

NEIBANK: EST analysis and bioinformatics for ocular genomics

JOURNAL

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

COMMENT

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 06 row: f column: 01

Seq primer: M13RPI reverse primer (ABI).

FEATURES

Location/Qualifiers

1..523

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="by06f01"

/tissue_type="Lens"

/dev_stage="Adult"

/lab_host="EMDH108"

/clone_lib="Human Lens cDNA (Un-normalized, unamplified):
BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pCMVSPORT6 vector was constructed at Life
Technologies, essentially following the protocols of the
SuperScript Plasmid System full details of which are
contained in the manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-pGACTAGTTCCTAGATCGGAGCGGCC(T)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 3 43e-98 Length: 523

Score: 893.00 Matches: 168

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 97.49% Indels: 0

DB:	2	Gaps:	0	FEATURES	source
US-10-657-740-1	(1-173) x BF726438 (1-523)			Location/Qualifiers	
QY	1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20			1. 588	
Db	18 ATGGAGGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCACG 77			/organism="Homo sapiens"	
QY	21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40			/mol_type="mRNA"	
Db	78 CGGTGTTTCGACCACTTTTCGGCGAGGCCCTTTTGTAGTATGATGACCTGCTGCCCTTCCTG 137			/db_xref="taxon:9606"	
QY	41 SerSerThrIleSerProTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60			/clone="UI-E-DW0-agk-p-02-0-UI"	
Db	138 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTTTCCGACACCTGCTGACCTCCGAC 197			/tissue_type="lens"	
QY	61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80			/dev_stages="adult"	
Db	198 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTCTGTCATCTTCTCGATGTGAAGCACTTC 257			/lab_host="DH10B (Life Technologies) (T1 phage resistant)"	
QY	81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100			/clone_lib="UI-E-DW0"	
Db	258 TCCCGGAGGACCTCACCGTGAAGGTGCAGGACGACTTTGTGAGATCCACGGAAGCAC 317			modified polylinker: Site 1: EcoR I; Site 2: Not I;	
QY	101 AsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120			UI-E-DW0 is a cDNA library containing the following	
Db	318 AACGAGCGCCAGGACGACCGCTACATTTCCGCTGAGTTCCACCGCGCTACCGCCTG 377			tissue(s): lens. The library was constructed according to	
QY	121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140			Bonaldo, Lennon and Soares, Genome Research, 6:791-806,	
Db	378 CGTCCACAGTGACGACGTCGGCCCTCTCTTCTCCCTGCTGCGATGCGATGTCGACC 437			1996. First strand cDNA synthesis was primed with an	
QY	141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160			oligo-dT primer containing a Not I site. Double stranded	
Db	438 TTCTGTGCCCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGCGGAGGACCATCCCC 497			cDNA was ligated to an EcoR I adaptor, digested with Not	
QY	161 ValSerArgGluGluLysProThr 168			I, and cloned directionally into pT73-Pac vector. The	
Db	498 GTGTGCGGGAGGAGAGCCACCC 521			oligonucleotide used to prime the synthesis of	
RESULT 31				first-strand cDNA contains a library tag sequence that is	
BM696799				located between the Not I site and the (dT)18 tail. The	
LOCUS				sequence tag for this library is CGATTAGCGA. This library	
DEFINITION				was created for the program, Gene Discovery in the Visual	
ACCESSION				System, supported by National Eye Institute (NEI)."	
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
COMMENT					
BM696799	588 bp	mRNA	linear	EST	28-FEB-2002
UI-E-DW0-agk-p-02-0-UI.r1	UI-E-DW0	Homo sapiens	cDNA clone		
UI-E-DW0-agk-p-02-0-UI 5', mRNA sequence.					
BM696799					
BM696799.1	GI:19010057				
EST.					
Homo sapiens (human)					
Homo sapiens					
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
1 (bases 1 to 588)					
Bonaldo,M.F., Lennon,G. and Soares,M.B.					
Normalization and subtraction: two approaches to facilitate gene					
discovery					
Genome Res. 6 (9), 791-806 (1996)					
97044477					
889548					
Contact: Soares, MB					
Coordinated Laboratory for Computational Genomics					
University of Iowa					
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA					
Tel: 319 335 8250					
Fax: 319 335 9565					
Email: bento-soares@uiowa.edu					
Tissue Procurement: Dr. Gregg Hageman					
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa					
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa					
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa					
Clone Distribution: Researchers may obtain clones from Research					
Genetics (www.resgen.com).					
Seq primer: M13 Reverse.					

ORIGIN

Alignment Scores:

Pred. No.: 4,02e-98 Length: 588
 Score: 893.00 Matches: 171
 Percent Similarity: 98.84% Conservatives: 0
 Best Local Similarity: 98.84% Mismatches: 2
 Query Match: 49.49% Indels: 1
 Ds: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM696799 (1-588)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 54 ATGGAGGTGACCATCCAGCACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCACG 113
 QY 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 114 CGGTGTTTCGACCACTTTTCGGCGAGGCCCTTTTGTAGTATGACCTGCTGCCCTTCCTG 173
 QY 41 SerSerThrIleSerProTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 174 TCGTCCACCATCAGCCCTTACTACCGCACGTCCCTTTCGGCACCGTGTGACCTCCGCG 233
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 234 ATCTCTGAGTTTCGATCCGACCGGACAAAGTTCTGTCATCTTCTCGATGTGAAGCACTTC 293
 QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 294 TCCCGGAGGACCTCACCGTGAAGGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 353
 QY 101 AsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 Db 354 AACGAGCGCCAGGACGACCGCTACATTTCCCGTGAAGTTCCACCGCGCTACCGCCTG 413
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 414 CCGTCCAAAGTGGACCACTGCGCCCTCTCTTCTGCTCCCTGCTGCGATGGCATGCTGACC 473
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 474 TTCTGTGCCCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGCGAGGACCATCCCC 533
 QY 161 ValSerArgGluGluLysProThrSerAlaProSer 173
 Db 534 GTGTGCGGGAGGAGAGCCACC-TCGGNTCCCTCGTCC 571

RESULT 32
 BM697101
 Locus
 DEFINITION UI-E-DW0-9gm-m-13-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 UI-E-DW0-9gm-m-13-0-UI 5', mRNA sequence.
 ACCESSION BM697101
 VERSION BM697101.1 GI:19010359
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 593)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.regen.com).
 Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
 1..593
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-9gm-m-13-0-UI"
 /tissue_type="adult"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT7T3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CGATTAGCGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:
 Pred. No.: 5,39e-98 Length: 593
 Score: 892.00 Matches: 168
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 97.38% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM697101 (1-593)

QY 26 PhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIleSer 45
 Db 61 TTTTTCGGCAGGGCTTTTGGATGATGACCTGCTGCCCTTCTCTGTCTCCACATCAGC 120
 QY 46 ProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluValArg 65
 Db 121 CCTTACTACCGCCAGTCCCTCTTCGCAACCGTGTGGACTCCGGCATCTCTGAGGTTCGA 180
 QY 66 SerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeu 85
 Db 181 TCCGACCCGGCAAGTTCGTCTCATCTTCGATGTGAAGACATTTCTCCCGGAGGACCTC 240
 QY 86 ThrValLysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAsp 105
 Db 241 ACCGTGAAGGTGCAGGACGACTTTGTGGAGATCACGGAAGACCAAGAGGCCGAGGAC 300
 QY 106 AspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAsp 125
 Db 301 GACCACGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTCGCTCCACCGTGGAC 360
 QY 126 GlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLys 145
 Db 361 CAGTCGGCCCTCTCTTGTCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGCCCCCAAG 420
 QY 146 IleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGlu 165
 Db 421 ATCCAGACTGCGCTGGATGCCACCCAGCGAGCGGACATCCCGTGTGCGGGAGGAG 480
 QY 166 LysProThrSerAlaProSerSer 173
 Db 481 AAGCCACCTCGGCTCCCTCGTCC 504

RESULT 33

BF726679
 Locus
 DEFINITION by10d02.yi Human Lens cDNA (Un-normalized, unamplified): BY Homo
 sapiens cDNA clone by10d02 5', mRNA sequence.
 ACCESSION BF726679
 VERSION BF726679.1 GI:12042590
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 558)
 AUTHORS Wistow,G.J., Bernsstein,S., Behal,A. and Smith,D.
 TITLE NEIBANK: EST analysis and bioinformatics for ocular genomics
 JOURNAL Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 10 row: d column: 02
 Seq primer: M13Rpl reverse primer (ABI).
 Location/Qualifiers
 1..558
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="by10d02"
 /tissue_type="Lens"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
 BY"

FEATURES

source
 1..558
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="by10d02"
 /tissue_type="Lens"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
 BY"
 /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
 from different adults (both approximately 40 years old)
 together yielded 20ug of total RNA and 150ng mRNA for cDNA
 library synthesis. A directionally cloned cDNA library in

QY 6 GlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGln 25
 Db 1 CAGCACCCCTGGTTCAAGCGCACCTTGGGGCCCTTCTACCCCGCGCTGTTCGACCAG 60

the pCMVSPORT6 vector was constructed at Life Technologies, essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor (5'-pGACTAGTTCTAGATCGGAGCGCGCC(T)15-3'). Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC).

ORIGIN

Alignment Scores:
 Pred. No.: 1.15e-97 Length: 558
 Score: 889.00 Matches: 168
 Percent Similarity: 99.41% Conservatives: 0
 Best Local Similarity: 99.41% Mismatches: 1
 Query Match: 97.05% Indels: 0
 DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF726679 (1-558)

Qy 5 lIeGlnHlaProTyrPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAap 24
 Db 2 ATCCAGCACCCCTGGTTCAAGCGCACCTTCTGAGGCTTCTACCCAGCGGCTGTTCGAC 61
 Qy 25 GlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrile 44
 Db 62 CAGTTTTTCGGGAGGCGCTTTTGGATGATGACCTGCTGCCCTTCTGTGCTCCACATC 121
 Qy 45 SerProTyrTyrArgGlnSerLeuPheArgThrValLeuAapSerGlyIleSerGluVal 64
 Db 122 AGCCCTACTACCGCCAGTCCCTTCCGCGACCGTGTGGATCGGCGCATCTCTGAGGTT 181
 Qy 65 ArgSerAspAspAspLysPheValIlePheLeuAapValIleHisPheSerProGluAap 84
 Db 182 CGATCCGACCGGAGCAAGTTGCTATCTTCTCGATGAGAGCATCTTCTCCCGGAGGAC 241
 Qy 85 LeuThrValIleValGlnAspPheValGluIleHisGlyLysHisAasnGluArgGln 104
 Db 242 CTCACCGTGAAGTGCAGGACGACTTTGTGGATGATCCAGGAAAGCAACAGCGCCAG 301
 Qy 105 AspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnVal 124
 Db 302 GACGACACCGGTACATTTCCGTCGATTCACCGCGCTACCGCTCGCGTCCACCGTG 361
 Qy 125 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 144
 Db 362 GACCAAGTCGGCCCTCTTGTCTCTCTGTCCTGCGATGCGATGCTGACCTTCTGTGGCCCC 421
 Qy 145 LysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGlu 164
 Db 422 AAGATCCAGACTGGCTGGATGCACCAACCGCGGAGCGGCGCATCTCCCGTGTGCGGGAG 481
 Qy 165 GluLysProThrSerAlaProSerSer 173
 Db 482 GAGAAGCCACCTCGGCTCCCTCGTCC 508

RESULT 34

BM696667
 LOCUS
 DEFINITION UI-E-DW0-agk-o-10-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 UI-E-DW0-agk-o-10-0-UI 5', mRNA sequence.
 BM696667
 BM696667.1 GI:19009925
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 558)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 889548

Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics

University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reagen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

source

1. 558
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agk-o-10-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT7T3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (GT)₁₈ tail. The
 sequence tag for this library is CGATTAGCGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:
 Pred. No.: 1.52e-97 Length: 558
 Score: 888.00 Matches: 171
 Percent Similarity: 98.84% Conservatives: 0
 Best Local Similarity: 98.84% Mismatches: 2
 Query Match: 96.94% Indels: 1
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM696667 (1-558)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 36 ATGGAGGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTTCTACCCAGC 95
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
 Db 96 CGGCTGTTTCGACCAAGTTTTTCGGCGAGGCGCTTTTGGATGATGACCTCTGCGCTTCTCG 155
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAapSerGly 60
 Db 156 TCGTCCACCATCAGCCCTACTACCGCGAGTCCCTCTTCCGACCGCTGTGACTCCGCGC 215
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAapValIleHisPhe 80
 Db 216 ATCTCTGAGGTTTCGATCCGACCGGAGCAAGTTCTGTCATCTTCTCGATGTGAAGCAC 275
 Qy 81 SerProGluAapLeuThrValLysValGlnAapAspPheValGluIleHisGlyLysHis 100

Db 276 TCCCGGAGGACCTCACCCTGAAGGTGAGGACGACCTTTGTGGAGATCCACGGAAGCAC 335
 Qy 101 AenGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
 Db 336 AACGAGCCGACGAGCACACCGCTACATTTCCCTGGAGTTCCACCGCGCTTACCGCGCTG 395
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 396 CGGTCCACAGTGGACAGTCCGGCTCTCTGTCTCCCTGTCTGCCGATGGATGCTGACC 455
 Qy 141 PheCysGlyProLyLeuGlnThrGlyLeuAspAlaThrHisAlaGluAlaIlePro 160
 Db 456 TTCTGTGGCCCCAAGATCAGACTGGCTGGATGCCACCCAGCGAGCGAGCCATCCCC 515
 Qy 161 ValSerArgGluGluValProThrSerAlaProSerSer 173
 Db 516 GTGTGCGGGAGGAGAGCCACC-TCGGCTCCCTCGTCC 553

RESULT 35
 LOCUS BM696489 508 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-DW0-agj-n-20-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 VERSION BM696489
 KEYWORDS BM696489.1 GI:19009747
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 508)
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@iowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..508
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-DW0-agj-n-20-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT7T3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The

sequence tag for this library is CGATTAGCGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:
 Pred. No.: 1.77e-97 Length: 508
 Score: 887.00 Matches: 167
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 96.83% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM696489 (1-508)

Qy 7 HisProTrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhe 26
 Db 1 CACCCCTGGTTCAAGCGCACCTCGGGCCCTTACCCACGCGGCTCTTCGACCATGTTT 60
 Qy 27 PheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIleSerPro 46
 Db 61 TTCGGCGAGGGCCCTTTTGTAGTATGACCTGCTGCCCTTCCCTGCTCCACCATCAGCCCC 120
 Qy 47 TyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSer 66
 Db 121 TACTACCGCCAGTCCCTCTCCGACCGCTGCGACCTCCGCGCATCTCTGAGGTTGATCC 180
 Qy 67 AspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThr 86
 Db 181 GACCGGGACAAAGTTTCGTCATCTTCCTGATGTGAAGCACTTCTCCCGGAGGACCTCACC 240
 Qy 87 ValLysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAsp 106
 Db 241 GTGAAGGTGCGAGGACGACTTTGTGGATCCACGAAAGCAACAGCGCCGACGAGGACAC 300
 Qy 107 HisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnValAspGln 126
 Db 301 CACGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCGCTCCGCTCAACGTTGACACAG 360
 Qy 127 SerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIle 146
 Db 361 TCGGCGCTCTCTGCTCCCTGCTGCGGATGGCATGTGACCTTCTGTGGCCCCCAAGATC 420
 Qy 147 GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluLys 166
 Db 421 CAGACTGGCTGGATGCCACCGAGCGAGCGCATCCCGTGTCTCGGGAGGAGAGAG 480
 Qy 167 ProThrSerAlaProSerSer 173
 Db 481 CCCACCTCGGCTCCCTCGTCC 501

RESULT 36
 LOCUS BM686748 663 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-CQ0-adq-b-11-0-UI.r1 UI-E-CQ0 Homo sapiens cDNA clone
 VERSION UI-E-CQ0-adq-b-11-0-UI 5', mRNA sequence.
 ACCESSION BM686748
 KEYWORDS BM686748.1 GI:19000006
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 663)
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: bento-soares@uiowa.edu
Tissue procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

source

1..663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CQ0-adj-b-11-0-UI"
/tissue_type="optic nerve"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker: Site 1: EcoR I; Site 2: Not I; UI-E-CQ0 is a cDNA library containing the following tissue(s): optic nerve. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCATTAACTG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 2,55e-97 Length: 663
Score: 887.00 Matches: 171
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 1
Query Match: 96.83% Indels: 1
DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM686748 (1-663)

Qy 2 AspValThrIleGlnHisProThrPheLysArgThrLeuGlyProPheTyProSerArg 21
Db 10 GACGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTTCTAGCGGCGCTTCTACCCAGCGG 69
Qy 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyArgPheLeuLeuProPheLeuSer 41
Db 70 CTGTTTCGACCAAGTTTTCGCGGAGGCGCTTTTTCGATGATGACCTGCTCCCTTCTCTGTCG 129
Qy 42 SerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGlyIle 61
Db 130 TCCACCATCACCCTTACTACCGCAGTCCCTCTTCGACCGTGTGATCTCGGCATC 189
Qy 62 SerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSer 81
Db 190 TCTGAGGTTCTGATCCGACGGGACAAAGTTCGTCTATCTTCTCTCGATGTGAAGCACTTCTCC 249
Qy 82 ProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHisAsn 101
Db 250 CCGGAGGACCTCACCGTGAAGGTGAGGACGACCTTTGTGGAGATCCACGGAAGCACAAC 309
Qy 102 GluArgGlnAspAspHisGlyTyTrIleSerArgGluPheHisArgAtgTyArgLeuPro 121
Db 310 GAGGCCACGAGGACGACGCGTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCGG 369
Qy 122 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 141

Db 370 TCCACGTTGGACCAAGTCCGCGCTCTCTTGTCTCCCTGTCTGCCGATGCTGACCTTC 429
Qy 142 CysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 161
Db 430 TGTGCCCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGGAGCGAGCCATCCCGGTG 489
Qy 162 SerArgGluGluLysProThrSerAlaProSerSer 173
Db 490 TCGCGGAGGAGAGAGCCACC-TCGGCTCCCTCGTCC 524

RESULT 37

BF727205

LOCUS

DEFINITION

sapiens cDNA clone by18b11 5', mRNA sequence.

ACCESSION

BF727205

VERSION

BF727205.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 503)

AUTHORS

Wistow, G.J., Bernstein, S., Behal, A. and Smith, D.

TITLE

NEIBANK: EST analysis and bioinformatics for ocular genomics

JOURNAL

Invest. Ophthalmol. Vis. Sci. 41 (2000) In press

COMMENT

Contact: Wistow G

Section on Molecular Structure and Function

National Eye Institute

6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452

Fax: 301 496 0078

Email: graeme@helix.nih.gov

Plate: 18 row: b column: 11

Seq primer: M13RP1 reverse primer (ABI).

Location/Qualifiers

1..503

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="by18b11"

/tissue_type="Lens"

/dev_stage="Adult"

/lab_host="EMDH10B"

/clone_lib="Human Lens cDNA (Un-normalized, unamplified): BY"

/note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses from different adults (both approximately 40 years old) together yielded 20ug of total RNA and 150ng mRNA for cDNA library synthesis. A directionally cloned cDNA library in the pCMVSPORT6 vector was constructed at Life Technologies essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTTTATGATCGGAGCGCGCC(T)15-3']. Not I/blunt end inserts were cloned into the Not I/EcoR V sites in the vector. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:

Pred. No.: 1.25e-96 Length: 503
Score: 880.00 Matches: 166
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 1
Query Match: 96.07% Indels: 0
DB: 2 Gaps: 0

US-10-657-740-1 (1-173) x BF727205 (1-503)

```
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 2 ATGGATGTGACCATCCAGACCCCTGGTTCAAGCGCACCTTGAGCCCTTCTACCCACG 61
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db 62 CGGCTGTTTCGACCACTTTTTCGGCGAGGCGCTTTTGTAGTATGACCTGCTGCCCTTCCTG 121
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 122 TCGTCCACATCAGCCCTTACTACCGCCAGTCCCTTTCGCGACCGTCTGGACTCCGGC 181
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db 182 ATCTCTGAGTTCGATCCGACCGGACAAAGTTCGTTCATCTTCTCGATGTGAGCACTTC 241
Qy 81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
Db 242 TCCCCAGAGGACCTCAGCGTGAAGGTGCAGGACGACTTGTGGAGATCCACGGAAGCAC 301
Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db 302 AACGAGCCCGAGGACGACCGCTACATTTCCCGTGAGTTCACCCGCGCTACCGCGCTG 361
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 362 CGGTCCACGTGGACCACTCGGCCCTCTTGTCTCCCTGTCTGCGATGGCATGCTGACC 421
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 422 TTCTGTGGCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGCGAGCGGACCATCCCC 481
Qy 161 ValSerArgGluGluLysPro 167
Db 482 GTGTGCGGGAGGAGAGCCCC 502

RESULT 38
BM696477 519 bp mRNA linear EST 28-FEB-2002
LOCUS UI-E-DW0-agj-1-20-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
DEFINITION UI-E-DW0-agj-1-20-0-UI 5', mRNA sequence.
ACCESSION BM696477
VERSION BM696477.1 GI:19009735
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 519)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reagen.com).
Seq primer: M13 Reverse
Location/Qualifiers
1..519
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db xref="taxon:9606"
/clone="UI-E-DW0-agj-1-20-0-UI"
/tissue_type="lens"
/dev_stage="adult"
/lab_hosts="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-DW0 is a cDNA library containing the following
tissue(s): lens. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGATTAGCGA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
```

ORIGIN

Alignment Scores:					
Pred. No.:	1.72e-96	Length:	519		
Score:	879.00	Matches:	166		
Percent Similarity:	99.40%	Conservative:	0		
Best Local Similarity:	99.40%	Mismatches:	1		
Query Match:	95.96%	Indels:	0		
DB:	4	Gaps:	0		

US-10-657-740-1 (1-173) x BM696477 (1-519)

Qy	7	HisProTyrPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhe	26
Db	3	CACCCCTGGTTCAAAGCGCACCTCGGGCCCTTACCCCGACCGCGCTTTCGACCACTTT	62
Qy	27	PheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIleSerPro	46
Db	63	TTCCGGCAGGGCCCTTTTGTAGTATGACCTGCTGCCCTTCTGTGTCACCATCAGCCCC	122
Qy	47	TyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSer	66
Db	123	TACTACCGCCAGTCCCTCTCCGACCGCTGCTGGACTCCCGCATCTCTGAGGTTCGATCC	182
Qy	67	AspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThr	86
Db	183	GACCGGACAAAGTTTCGTCATCTTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACC	242
Qy	87	ValLysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAsp	106
Db	243	GTGAAGGTGCAGGACGACTTGTGGAGATCCACGGAAGCACACAGCGCGCAGGACGAC	302
Qy	107	HisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGln	126
Db	303	CACGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTCCGCTCCCAACGTGGACCAG	362
Qy	127	SerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIle	146
Db	363	TCGGCCCTCTCTTGTCTCCCTGTCTGCGCATGGCATGCTGACCTTCTGTGGGCCCAAGATC	422
Qy	147	GlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLys	166
Db	423	CAGACTGGCTGGATGCCACCGCGAGCGAGGCCATCCCCGTGTGTCGGGAGGAGGAAA	482
Qy	167	ProThrSerAlaProSerSer	173
Db	483	GCCACCTCGGCTCCCTCGTCC	503

RESULT 39
BM686206
LOCUS UI-E-C11-aay-d-12-0-UI.r1 UI-E-C11 Homo sapiens cDNA clone
DEFINITION UI-E-C11-aay-d-12-0-UI.r1 UI-E-C11 Homo sapiens cDNA clone

```

ACCESSION      BM686206
VERSION        BM686206.1  GI:18996102
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheraia; Primates; Catarrhini; Hominiidae; Homo.
TITLE         Bonaldo,M.F., Lennon,G. and Soares,M.B.
              Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL        Genome Res. 6 (9), 791-806 (1996)
MEDLINE        9704477
PUBMED         8889548
COMMENT        Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: bento-soares@uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.regen.com).
              Seq primer: M13 Reverse.
              Location/Qualifiers
                1..536
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="UI-E-C11-aay-d-12-0-UI"
                  /tissue_type="RPE and Choroid"
                  /dev_stage="adult"
                  /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                  /clone_lib="UI-E-C11"
                  /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                  modified polylinker; Site 1: EcoR I; Site 2: Not I;
                  UI-E-C11 is a normalized cDNA library containing the
                  following tissues: RPE and Choroid. The library was
                  constructed according to Bonaldo, Lennon and Soares,
                  Genome Research, 6:791-806, 1996. First strand cDNA
                  synthesis was primed with an oligo-dT primer containing a
                  Not I site. Double stranded cDNA was ligated to an EcoR I
                  adaptor, digested with Not I, and cloned directionally
                  into pT7T3-Pac vector. The oligonucleotide used to prime
                  the synthesis of first-strand cDNA contains a library tag
                  sequence that is located between the Not I site and the
                  (dT)18 tail. The sequence tag for this library is ACCTA.
                  This library was created for the program, Gene Discovery
                  in the Visual System, supported by National Eye Institute
                  (NEI)."
```

ORIGIN

Alignment Scores:

Pred. No.:	3,166-96	Length:	536
Score:	877.00	Matches:	165
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	1
Query Match:	95.74%	Indels:	0
DB:	4	Gaps:	0

US-10-657-740-1 (1-173) x BM686206 (1-536)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyVProPheTyrProSer 20
 |||||
 Db 39 ATGGATGTGACCATCCAGACCCCTGGTTTCAGCGCACCTGGGGCCCTTTACCCGAGC 98
 |||||
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuPheLeu 40
 |||||

Db 99 CGGTGTTTCGACCAAGTTCCTTCGGCGAGGGCCCTTTTGTAGTATGACCTGCTGCTTCCTG 158
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 |||||
 Db 159 TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCCGACCGTCTGACTCCGCGC 218
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPhe 80
 |||||
 Db 219 ATCTCTGAGGTTCGATCCGACCGGACAAAGTTCGTTCATCTTCTCTCGATGTGAAGCACTTC 278
 Qy 81 SerProGluAspLeuThrValIleValGlnAspAspPheValGluIleHisGlyLysHis 100
 |||||
 Db 279 TCCCGGAGGACCTCACCGTGAAGTGCAGACGACCTTTGTGGAGATCCACGAAAGCAC 338
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
 |||||
 Db 339 AACGAGCGCCAGGACGACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTG 398
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 |||||
 Db 399 CCGTCCAAACGTGGACCAAGTTCGGGCTCTCTTGTCTCCCTGTCTGCGATGGCATGCTGACC 458
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 |||||
 Db 459 TTCTGTGGCCCAAGATCCAGACTGGCTGGATGCCACCGCGCGAGGAGCCATCCCC 518
 Qy 161 ValSerArgGluGluLys 166
 |||||
 Db 519 GTGTCCGCGGAGGAGAG 536

RESULT 40
 BQ637175
 LOCUS
 DEFINITION he06f05.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
 Homo sapiens cDNA clone he06f05 5', mRNA sequence.
 ACCESSION BQ637175
 VERSION EST.
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheraia; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 542)
 AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A.,
 Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.
 TITLE Expressed sequence tag analysis of human retina for the NEIBank
 Project: Retbindin, an abundant, novel retinal cDNA and alternative
 splicing of other retina-preferred gene transcripts
 JOURNAL Mol. Vis. 8 (4), 196-204 (2002)
 MEDLINE 22103461
 PUBMED 12107411
 COMMENT Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 06 row: f column: 05
 Seq primer: M13RP1 reverse primer (ABI).
 Location/Qualifiers
 1..542
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="he06f05"
 /tissue_type="Retina"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Human Retina cDNA (Un-normalized,
 unamplified): hd/he"
 /note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
 was dissected from two 80 year old donors with no observed

FEATURES
 source

eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor
 [5'-pGACTAGTTCTAGATCGGAGCGGCCGCC(T)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
 Pred. No.: 4,248-96 Length: 542
 Score: 876.00 Matches: 165
 Percent Similarity: 100.00% Conservatve: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 95.63% Indels: 0
 DB: 5 Gaps: 0

US-10-657-740-1 (1-173) x BQ637175 (1-542)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
 Db ATGGATGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGCCCTTCTACCCACG 106
 Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 107 CGGTGTTCGACAGCTTTTCGGCGAGGCGCTTTTGGATATGACCTGCTGCCCTTCCTG 166
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 167 TGTCTCCACCATCAGCCCTTACTACCGCAGTCCCTCTTCGCGACCGTCTGGACTCCGCG 226
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 227 ATCTCTGAGTTCGATCCGCGGACCAAGTTCGTCATCTTCCTCGATGTGAGCAGCTTC 286
 Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 287 TCCCGGAGGACCTCACCCTGAAGGTGAGGACGACTTTGTGGAGATCCACGGAAAGCAC 346
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 Db 347 AACGAGCGCCAGGACGACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCGCTG 406
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 407 CGGTCCACGTGGACAGTGGCCCTCTCTTGCTCCCTGCTGCGATGGCATGTGACC 466
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 467 TTCTGTGGCCCAAGATCCAGACTGGCTGGATGCCACCCAGCGCGAGCGGACCATCCCC 526
 Qy 161 ValSerArgGluGlu 165
 Db 527 GTGTGCGGAGGAG 541

RESULT 41

BM696853
 LOCUS BM696853 498 bp mRNA linear EST 28-FEB-2002
 DEFINITION UI-E-DW0-agl-k-14-0-UI.r1 UI-E-DW0 Homo sapiens cDNA clone
 UI-E-DW0-agl-k-14-0-UI 5', mRNA sequence.

ACCESSION BM696853

VERSION BM696853.1

KEYWORDS EST: GI:19010111

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 498)

AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 8889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES

source

1..498
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clones="UI-E-DW0-agl-k-14-0-UI"
 /tissue_type="lens"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-DW0"
 /notes="Organ; eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DW0 is a cDNA library containing the following
 tissue(s): lens. The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was ligated to an EcoR I adaptor, digested with Not
 I, and cloned directionally into pT7T3-Pac vector. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is CGATTAGCGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 8,76e-96 Length: 498
 Score: 873.00 Matches: 164
 Percent Similarity: 99.39% Conservatve: 0
 Best Local Similarity: 99.39% Mismatches: 1
 Query Match: 95.31% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BM696853 (1-498)

Qy 5 IleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAsp 24
 Db 2 ATCCAGCACCCCTGGTTCAAGCGCACCCCTTCTACCCGACCGCGCTGTTCGAC 61
 Qy 25 GlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIle 44
 Db 62 CAGTTTTTCGGCGAGGCGCTTTTGGATGACCTGCTGCCCTTCTGTGCTCCACCATC 121
 Qy 45 SerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluVal 64
 Db 122 AGCCCTTACTACCGCAGTCCCTCTTCGCGACCGTGTGGACTCCGCGCATCTCTGAGGTT 181
 Qy 65 ArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAsp 84
 Db 182 CGATCCGACCGGGCAAGTTGCTCATCTTCCTCGATGTGAAGCACCTTCTCCCCGAGGAC 241

Qy 85 LeuThrValysValGlnAspPheValGluIleHisGlyLysHisAsnGluArgGln 104
 Db 242 CTCACCGTGAAGGTGCAGGACGACTTTGTGAGATCCACGGAACACCAAGGAGCCGAC 301

Qy 105 AspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnVal 124
 Db 302 GACGACACGAGGTACATTTCCGCGAGTTCACCGCGCTACCGCTCCGCTCCACAGTG 361

Qy 125 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 144
 Db 362 GACCAGTCGGNCCTCTCTTGTCTCCCTGTCTGCGATGCGATGCTGACCTTCTGTGGCCCC 421

Qy 145 LysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGlu 164
 Db 422 AAGATCCAGACTGGCGTGGATGCCACCCAGCCGAGCGAGCATCCCGCTGTCGGGGAG 481

Qy 165 GluLysProThrSer 169
 Db 482 GAGAGGCCACCTCG 496

RESULT 42
 CF732528
 LOCUS
 DEFINITION UI-M-HAO-cjw-j-03-0-UI.r1 NIH_BMAP_HAO Mus musculus cDNA clone
 IMAGE:30551930 5', mRNA sequence.
 CF732528
 VERSION
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 688)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5
 Location/Qualifiers
 1. 688
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30551930"
 /dev_stage="embryo 12.5,13.5,14.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_HAO"
 /note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is TTATTGAAGT. This library was created for the University
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

FEATURES
 source

ORIGIN
 Alignment Scores:
 Pred. No.: 1.36e-95 Length: 688
 Score: 873.00 Matches: 164
 Percent Similarity: 97.69% Conservative: 5
 Best Local Similarity: 94.80% Mismatches: 4
 Query Match: 95.31% Indels: 0
 DB: 7 Gaps: 0

US-10-657-740-1 (1-173) x CF732528 (1-688)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 50 ATGGACGTTCACCATTCAGCATCTCTTGGTTCAAGCGTGCCTCTGGGGCCCTCTACCCCCAGC 109

Qy 21 ArgLeuPheAspGlnPhePheGlyGluLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 110 CGACTGTTTCACCAAGTTCTTCGGCGAGGGCCCTTTTGGTACGACCTGCTGCCCTCTCTG 169

Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 170 TCTTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCGCGACTGTCTGGACTCGGGC 229

Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 230 ATCTCTGAGGTCCGATCTGACCGGACAAAGTTTGTCTATCTTCTTGGACGTGAAGCACTTC 289

Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 290 TCTCTGAGGACCTTCACCGTGAAGTACTGGAGATTTTGGAGATTTCACGGANACAC 349

Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 Db 350 AACGAGAGCGAGGATGACCATGGCTACATTTCCCGTGATTTCCCGTCCGTACCGCTCG 409

Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 410 CTTTCCAAATGGACACAGTCGCGCTCTCTCTGCTCCCTGTCTGGGATGGATGTGTGACC 469

Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 470 TTCTCTGCGCCCAAGTCCAGTCCGGTTGGATGCTGTGCACGAGAGGGCCATTCCT 529

Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 530 GTGTACGGGAGGAGAACCCAGCTCTGCACCTCTGCTCC 568

RESULT 43
 AY419531
 LOCUS
 DEFINITION Mus musculus CRYAA gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY419531
 VERSION AY419531.1 GI:39775488
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 522)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 Science 302 (5652), 1960-1963 (2003)

TITLE
 JOURNAL
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 522)
 Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
 Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT These sequences were made by sequencing genomic exons and ordering
 them based on alignment.

FEATURES
 source
 1..522
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..522
 /gene="CRYAA"
 /locus_tag="HMC6921"

ORIGIN

Alignment Scores:
 Pred. No.: 1,646-95 Length: 522
 Score: 871.00 Matches: 164
 Percent Similarity: 97.69% Conservative: 5
 Best Local Similarity: 94.80% Mismatches: 4
 Query Match: 95.09% Indels: 0
 DB: 9 Gaps: 0

US-10-657-740-1 (1-173) x AY419531 (1-522)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 1 ATGGACGTCACCATTCACGATCCTTGGTTCAAGCGTGCCTGGGGCCCTTCTACCCACG 60
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 61 CGACTGTTTCACCATTCCTTCGGCGAGGGCCCTTTTGTAGTACGACCTGCTGCCCTTCTG 120
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 121 TCTTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCCGACATGTCTGGACTCGGGC 180
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 181 ATCTCTGAGGTCGACATCTGACCGGACAAAGTTTGTTCATCTTCTGGACGTGAAGCACTTC 240
 Qy 81 SerProGluAspLeuThrValLysValGlnAspPhePheValGluIleHisGlyLysHis 100
 Db 241 TCTCTCTGAGGACCTCACCCTGTAAGTACTGAGGATTTTGTGGAGATTACCGGCACAC 300
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
 Db 301 AACGAGAGGCGAGGATGACATGGCTACATTTCCCGTGAATTTACCGCTCGCTACCGTCTG 360
 Qy 121 ProSerAnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 361 CTTTCCAATGTGGACCACTCGGCCCTCTCTGCTCCCTCTCTGCGGATGGCATGCTGACC 420
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 421 TTCTCTGCCCCCAAGCTCAGTCCGGTTTGATGCTGCCACAGCGAGAGGGCCATTCTCT 480
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 481 GTGTCACGGGAGGAGAAACCCAGCTCTGCACCCCTCGTCC 519

RESULT 44
 CK628252
 LOCUS CK628252 552 bp mRNA linear EST 26-JAN-2004
 DEFINITION ip09h08.y2 Mouse whole eye, unamplified: io/ip Mus musculus cDNA
 clone ip09h08 5', mRNA sequence.
 ACCESSION CK628252
 VERSION CK628252.1 GI:41349138
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 552)
 Wistow,G. and Tomarev,S.
 Expressed sequence tag analysis of mouse whole eye
 Unpublished (2004)
 Contact: Wistow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 09 row: h column: 08
 Seq primer: M13RP1 reverse primer (ABI).
 Location/Qualifiers
 1..552
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57Bl6J"
 /db_xref="taxon:10090"
 /clone="ip09h08"
 /tissue_type="Whole eye"
 /dev_stage="Adult"
 /lab_host="EMDH108"
 /clone_lib="Mouse Whole eye, unamplified: io/ip"
 /note="Organ: Eye; Vector: pSport1; Approximately 1mg
 total RNA was extracted from 200 adult mouse whole eyes. A
 directionally cloned cDNA library in the pSPORT1 vector
 (Invitrogen) was constructed at Bioserve Biotechnology
 (Laurel MD) essentially following the protocols of the
 SuperScript Plasmid System full details of which are
 contained in the manufacturer's instruction manual
 (http://www.lifetech.com/). First strand synthesis was
 carried out using a Not I primer-adaptor
 [5'-pGACTAGTCTAGTCGAGCGCGCC(T)15-3']. cDNA was
 cloned in Not I/Sal I sites. EST analysis was performed on
 the unamplified library at the NIH Intramural Sequencing
 Center (NISC)."

FEATURES

source

1..552
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57Bl6J"
 /db_xref="taxon:10090"
 /clone="ip09h08"
 /tissue_type="Whole eye"
 /dev_stage="Adult"
 /lab_host="EMDH108"
 /clone_lib="Mouse Whole eye, unamplified: io/ip"
 /note="Organ: Eye; Vector: pSport1; Approximately 1mg
 total RNA was extracted from 200 adult mouse whole eyes. A
 directionally cloned cDNA library in the pSPORT1 vector
 (Invitrogen) was constructed at Bioserve Biotechnology
 (Laurel MD) essentially following the protocols of the
 SuperScript Plasmid System full details of which are
 contained in the manufacturer's instruction manual
 (http://www.lifetech.com/). First strand synthesis was
 carried out using a Not I primer-adaptor
 [5'-pGACTAGTCTAGTCGAGCGCGCC(T)15-3']. cDNA was
 cloned in Not I/Sal I sites. EST analysis was performed on
 the unamplified library at the NIH Intramural Sequencing
 Center (NISC)."

ORIGIN

Alignment Scores:
 Pred. No.: 1,776-95 Length: 552
 Score: 871.00 Matches: 164
 Percent Similarity: 97.69% Conservative: 5
 Best Local Similarity: 94.80% Mismatches: 4
 Query Match: 95.09% Indels: 0
 DB: 7 Gaps: 0

US-10-657-740-1 (1-173) x CK628252 (1-552)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 5 ATGGACGTCACCATTCACGATCCTTGGTTCAAGCGTGCCTGGGGCCCTTCTACCCACG 64
 Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 65 CGACTGTTTCACCATTCCTTCGGCGAGGGCCCTTTTGTAGTACGACCTGCTGCCCTTCTG 124
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 125 TCTTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCCGACATGTCTGGACTCGGGC 184
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 185 ATCTCTGAGGTCGACATCTGACCGGACAAAGTTTGTTCATCTTCTGGACGTGAAGCACTTC 244
 Qy 81 SerProGluAspLeuThrValLysValGlnAspPhePheValGluIleHisGlyLysHis 100
 Db 245 TCTCTCTGAGGACCTCACCCTGAAAGTACTGGAGGATTTTGTGGAGATTACCGGCACAC 304
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
 Db 305 AACGAGAGGCGAGGATGACCATGGCTACATTTCCCGTGAATTTACCGCTCGCTACCGTCTG 364

```

QY      121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
      |||
      365 CCTTCCAATGTGACCAAGTCCGCGCTCTCTGCTCCCTGTCTGCGATGCGATGCTGACC 424
QY      141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
      |||
      425 TTCTCTGGCCCCCAGGTCCAGTCCGCTTTGGAGTCTGGCCACAGCGAGAGGCGCATTCCT 484
QY      161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
      |||
      485 GTGTACGGGAGGAGAAACCCAGCTCTGCACCTCGTCC 523

```

```

RESULT 45
CK627503
LOCUS      io04g04.y1 Mouse whole eye, mRNA linear EST 26-JAN-2004
DEFINITION clone io04g04 5', mRNA sequence.
ACCESSION CK627503
VERSION    CK627503.1 GI:41348389
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)

```

```

ORGANISM
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 557)
Wistow.G. and Tomarev,S.
Expressed sequence tag analysis of mouse whole eye
JOURNAL
COMMENT
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 04 row: 9 column: 04
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1..557
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57Bl6J"
/db_xref="taxon:10090"
/clone="io04g04"
/tissue_type="Whole eye"
/dev_stage="Adult"
/lab_host="EMDH10B"

```

```

FEATURES
source
1..557
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57Bl6J"
/db_xref="taxon:10090"
/clone="io04g04"
/tissue_type="Whole eye"
/dev_stage="Adult"
/lab_host="EMDH10B"

```

```

Alignment Scores:
Pred. No.:      1.79e-95      Length:      557
Score:          871.00      Matches:    164
Percent Similarity: 97.69%      Conservative: 5
Best Local Similarity: 94.80%      Mismatches: 4
Query Match:     95.09%      Indels:    0
DB:              7          Gaps:    0

US-10-657-740-1 (1-173) x CK627503 (1-557)

```

ORIGIN

```

Alignment Scores:
Pred. No.:      1.79e-95      Length:      557
Score:          871.00      Matches:    164
Percent Similarity: 97.69%      Conservative: 5
Best Local Similarity: 94.80%      Mismatches: 4
Query Match:     95.09%      Indels:    0
DB:              7          Gaps:    0

```

US-10-657-740-1 (1-173) x CK627503 (1-557)

```

QY      1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
      |||
      11 ATGGACGTCAACATTCAGCGTGCCTTCAAGCGTGCCTTGGGGCCCTTCTACCCGAC 70
QY      21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
      |||
      71 CGACTGTGTGACCAAGTCTTCGGCGAGGGCCCTTTTGTAGTAGACCTGTGCTTCCCTG 130
QY      41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
      |||
      131 TCTTCCACCATCAGCCCTTACTACCGCAGTCCCTTTCGGCAGCTGTGTGGACTCGGGC 190
QY      61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
      |||
      191 ATCTCTGAGTCCGATCTGACCGGACCAAGTTTGTTCATCTTCTGGACGTGAAGCACTTC 250
QY      81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
      |||
      251 TCTCTGAGGACCTCACCGTGAAGGTACTGGAGGATTTTGTGGAGATTTCACGGCAACAC 310
QY      101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
      |||
      311 AACGAGAGCGAGGATGACCATGGCTACATTTCCCGTGAATTTCCCGTCCGCTACCGTCTG 370
QY      121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
      |||
      371 CCTTCCAATGTGACCAAGTCCGCGCTCTCTCTGCTGCGATGCGATGCGATGCGTACC 430
QY      141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
      |||
      431 TTCTCTGGCCCCCAGGTCCAGTCCGCTTTGGATGTGGCCACAGCGAGAGGCGCATTCCT 490
QY      161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
      |||
      491 GTGTACGGGAGGAGAAACCCAGCTCTGCACCTCGTCC 529

```

RESULT 46

```

CK627631
LOCUS      io06f03.y1 Mouse whole eye, mRNA linear EST 26-JAN-2004
DEFINITION clone io06f03 5', mRNA sequence.
ACCESSION CK627631
VERSION    CK627631.1 GI:41348517
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)

```

ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 558)
Wistow.G. and Tomarev,S.
Expressed sequence tag analysis of mouse whole eye
JOURNAL
COMMENT
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 06 row: f column: 03
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers
1..558
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57Bl6J"
/db_xref="taxon:10090"
/clone="io06f03"
/tissue_type="Whole eye"
/dev_stage="Adult"
/lab_host="EMDH10B"

```

FEATURES

```

source
1..558
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57Bl6J"
/db_xref="taxon:10090"
/clone="io06f03"
/tissue_type="Whole eye"
/dev_stage="Adult"
/lab_host="EMDH10B"

```

/note="Organ: Eye; Vector: pSport1; Approximately 1mg total RNA was extracted from 200 adult mouse whole eyes. A directionally cloned cDNA library in the pSPORT1 vector (Invitrogen) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCCCTTT15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:
 Pred. No.: 1,79e-95 Length: 558
 Score: 871.00 Matches: 164
 Percent Similarity: 97.69% Conservative: 5
 Best Local Similarity: 94.80% Mismatches: 4
 Query Match: 95.09% Indels: 0
 DB: 7 Gaps: 0

US-10-657-740-1 (1-173) x CK627631 (1-558)

QY 1 MetAspValThrIleGlnHisProTtpPheLysArgThrLeuGlyProPheTyrProSer 20
 DB 11 ATGGAGCTCACCATTTCAGCATCTCTGGTTCAGCGTGCCCTGGGCCCCCTTCTACCCACG 70
 QY 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 DB 71 CGACTGTTTCGACGATCTTCGGCGAGGCGCTTTTGTAGTACGACCTGCTGCCCTTCCTG 130
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 DB 131 TCTTCACCATCAGCCCTTACTACGCCAGTCCCTCTTCCGACACTGTGCTGACCTGGGC 190
 QY 61 IleSerGluValArgSerAspArgLysPheValIlePheLeuAspVallyHisPhe 80
 DB 191 ATCTCTGAGGTCCGATCTGACCGGACAGATTGTTCATCTTCTGGACGTGAAGCACTTC 250
 QY 81 SerProGluAspLeuThrVallyHisValGlnAspPheValGluIleHisGlyLysHis 100
 DB 251 TCTCTCTGAGGACTTACCCTGAAGGTACTGGAGGATTTTGTGGAGATTTCACGGCAACAC 310
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 DB 311 AACGAGAGCAGGATGACCATGGCTACATTTCCCGTGAATTTACCGCTGCTACCGTCTG 370
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 DB 371 CTTTCCATGTGACACAGTCCGCCCTCTCTGCTCTCTGCTGCGGATGGCATGTGACC 430
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaLeuPro 160
 DB 431 TTTCTCTGGCCCCAAGTCCAGTCCGGTTTGGATGCTGGCCACAGCGAGAGGCCCATTCCT 490
 QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 DB 491 GTGTACGGGAGGAGAAACCCAGCTCTGCACCCCTCGTCC 529

RESULT 47

LOCUS CK627600 566 bp mRNA linear EST 26-JAN-2004
 DEFINITION io06c02.y1 Mouse Whole eye, unamplified: io/ip Mus musculus cDNA clone io06c02 5', mRNA sequence.
 ACCESSION CK627600
 VERSION CK627600.1 GI:41348486
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 566)
 Wislow,G. and Tomarev,S.
 Expressed sequence tag analysis of mouse whole eye
 JOURNAL Unpublished (2004)
 COMMENT Contact: Wislow G
 Section on Molecular Structure and Function
 National Eye Institute
 6/331, NIH, Bethesda, MD 20892-2740, USA
 Tel: 301 402 3452
 Fax: 301 496 0078
 Email: graeme@helix.nih.gov
 Plate: 06 row: c column: 02
 Seq primer: M13RP1 reverse primer (ABI).
 Location/Qualifiers
 1..566
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57Bl6J"
 /db_xref="taxon:10090"
 /clone="io06c02"
 /tissue_type="Whole eye"
 /dev_stage="Adult"
 /lab_host="EMDH10B"
 /clone_lib="Mouse Whole eye, unamplified: io/ip"
 /note="Organ: Eye; Vector: pSport1; Approximately 1mg total RNA was extracted from 200 adult mouse whole eyes. A directionally cloned cDNA library in the pSPORT1 vector (Invitrogen) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCCCTTT15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

FEATURES

source

ORIGIN

Alignment Scores:
 Pred. No.: 1,83e-95 Length: 566
 Score: 871.00 Matches: 164
 Percent Similarity: 97.69% Conservative: 5
 Best Local Similarity: 94.80% Mismatches: 4
 Query Match: 95.09% Indels: 0
 DB: 7 Gaps: 0

US-10-657-740-1 (1-173) x CK627600 (1-566)

QY 1 MetAspValThrIleGlnHisProTtpPheLysArgThrLeuGlyProPheTyrProSer 20
 DB 25 ATGGAGCTCACCATTTCAGCATCTCTGGTTCAGCGTGCCCTGGGCCCCCTTCTACCCACG 84
 QY 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 DB 85 CGACTGTTTCGACGATCTTCGGCGAGGCGCTTTTGTAGTACGACCTGCTGCCCTTCCTG 144
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 DB 145 TCTTCACCATCAGCCCTTACTACGCCAGTCCCTCTTCCGACACTGTGCTGACCTGGGC 204
 QY 61 IleSerGluValArgSerAspArgLysPheValIlePheLeuAspVallyHisPhe 80
 DB 205 ATCTCTGAGGTCCGATCTGACCGGACAGATTGTTCATCTTCTGGACGTGAAGCACTTC 264
 QY 81 SerProGluAspLeuThrVallyHisValGlnAspPheValGluIleHisGlyLysHis 100
 DB 265 TCTCTCTGAGGACTTACCCTGAAGGTACTGGAGGATTTTGTGGAGATTTCACGGCAACAC 324
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 DB 325 AACGAGAGCAGGATGACCATGGCTACATTTCCCGTGAATTTACCGCTGCTACCGCTG 384

total RNA was extracted from 200 adult mouse whole eyes. A directionally cloned cDNA library in the pSPORI vector (Invitrogen) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript plasmid system full details of which are contained in the manufacturer's instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACATGTTTCATCGAGCGGCCCT(T)15-3']. cDNA was cloned in Not I/Sal I sites. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:	1.87e-95	Length:	574
Pred. No.:	871.00	Matches:	164
Score:	97.69%	Conservative:	5
Percent Similarity:	97.69%	Mismatches:	4
Best Local Similarity:	98.80%	Indels:	0
Query Match:	95.09%	Gaps:	0
DB:	7		

US-10-657-740-1 (1-173) x CK627249 (1-574)

QY	1	MetAspValThrIleGlnHisProTyrPheIysArgThrLeuGlyProPheTyrProSer	20
DB	44	ATGGAGCGTCACCAATTACGATCTTTGGTTCAAGCGTGCCCTGGGGCCCTTTCTACCCGAG	103
QY	21	ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu	40
DB	104	CGACTGTTTCGACAGCTCTTTCGGCGAGGGCCCTTTTTCGATACGACCTGCTGCCCTTCCTCG	163
QY	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60
DB	164	TCTTCACCATCAGCGCCCTACTACCGCCAGTCCCTCTTCGCGCACTGTGCTGGAGCTCGGGC	223
QY	61	IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe	80
DB	224	ATCTCTGAGTCCGATCTGACCGGGACCAAGTTTGTCACTCTTCGGACGTGAGACACTTC	283
QY	81	SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis	100
DB	284	TCTCTCTGGGACCTCACCGTGAAGTACTCGAGGATTTTGTGGAGATTTACCGGCAACAC	343
QY	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
DB	344	AACGAGAGCGAGGATGACATGGCTACATTTCCGTGATTTTCACTCGTCCGTACCGCTCTCG	403
QY	121	ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	140
DB	404	CCTTCCAATGTGGACAGTCCGCGCCCTCTCTCTGCTCCCTGTCTGGGATGGCACTGCTGACC	463
QY	141	PheCysGlyProLysIleGluThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro	160
DB	464	TTTCTCTGGCCCCAAGGTCAGTCCGGTTTGATGCTGGCCACACGAGAGGGCCATTCTCT	523
QY	161	ValSerArgGluGlyLeuProThrSerAlaProSerSer	173
DB	524	GTGTCACGGAGAGAAACCCAGCTCTGCACCTCGTCC	562

RESULT 50	CK628200	579 bp	linear	EST 26-JAN-2004
LOCUS	CK628200			
DEFINITION	ip09b11.v2 Mouse Whole eye, unamplified: io/ip Mus musculus cDNA clone ip09b11 5', mRNA sequence.			
ACCESSION	CK628200			
VERSION	CK628200.1	GI:41349086		
KEYWORDS	EST.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 579)			

[illegible]

Wistow, G. and Tomarev, S.
Expressed sequence tag analysis of mouse whole eye
Unpublished (2004)
Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 09 row: b column: 11
Seq primer: M13RP1 reverse primer (ABI).

FEATURES

source

```
1. 579
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL6J"
/db_xref="taxon:10090"
/clone="ip09b11"
/tissue_type="Whole eye"
/dev_stage="Adult"
/lab_host="WMDH108"
```

ORIGIN

Alignment Scores:			
Pred. No.:	1.89e-95	Length:	579
Score:	871.00	Matches:	164
Percent Similarity:	97.69%	Conservative:	5
Best Local Similarity:	94.80%	Mismatches:	4
Query Match:	95.09%	Indels:	0
DB:	7	Gaps:	0
US-10-657-740-1 (1-173) x CK628200 (1-579)			
Qy	1	MetAspValThrIleGlnHisProTyrPheIysArgThrLeuGlyProPheTyrProSer	20
Db	55	ATGGACGTCAACATTCAGCATCCTGGTTCAAGCGTCCCTGGGGCCCTTCTACCCGAC	114
Qy	21	ArgLeuPheAspGlnPheGlyGluGlyLeuPheGlyIlePheLeuPheLeuProPheLeu	40
Db	115	CGACTGTTCGACCACTTCTCGGGAGGGCTTTTGAGTACGACCTGCTGCCCTTCCG	174
Qy	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60
Db	175	TCTTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCGCGACTGTGTGGACTCGGC	234
Qy	61	IleSerGluValArgSerAspArgAspIysPheValIlePheLeuAspValIysHisPhe	80
Db	235	ATCTCTGAGGTCCGATCTGACCGGACAAGTTTGTCATCTCTTGACGATGAAGCACTT	294
Qy	81	SerProGluAspLeuThrValIysValGlnAspPheValGluIleHisGlyLysHis	100
Db	295	TCTCTCTGAGGACCTCACCCTGAAGGTACTGGAGGATTTTGTGGAGATTTCACGGCAAAC	354
Qy	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
Db	355	AACGAGAGCGCAGGATGACCATGGGTACATTTCCCGTGAATTTCCCGTCCGTACCGGTCTG	414
Qy	121	ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr	140

```
Db 415 CCTTCAATGTGACACAGTCGCCCTCTCCTGCTCCCTGTCTGCGGATGGCATGTGACC 474
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 475 TTCTCTGGCCCCCAAGTCCAGTCCGGTTTGGATGCTGGCCACAGCGAGAGGGCCATTCTT 534
Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 535 GTGTCACGGGAGGAGAAACCCAGCTCTGCACCCCTCGTCC 573
```

Search completed: May 30, 2005, 08:47:54
Job time : 3273.53 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2005, 23:18:18 ; Search time 4210.45 Seconds

(without alignments)
1990.942 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVTIQHFWKRTLGPFYPS.....HAERAIPVSRBEKTPSAPSS 173

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10657740/runat_27052005_165253_3315/app_query.fasta_1.590
-DB=GenEmbl -QWTF=frascap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=150
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10657740 -CGCN 1.1 4362 @runat_27052005_165253_3315 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.hcg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	100.0	519	9	CR407691 Homo sapi
2	916	100.0	614	9	BC069528 Homo sapi
3	916	100.0	741	9	U66584 Human alpha
4	916	100.0	1112	9	HSU05569

5	916	100.0	1114	6	CQ731849	CQ731849 Sequence
6	879	96.0	543	4	OCRCYAA	X95382 O.cuniculus
7	871	95.1	1056	10	RNU47922	U47922 Rattus norv
8	869	94.9	793	4	BOVCRYA	M26142 Bovine alph
9	849.5	92.7	796	10	MMU310308	AJ310308 Mus muscu
10	849.5	92.7	1271	10	RNU47921	U47921 Rattus norv
11	810	88.4	1025	10	MUSCRYAAB	J00376 Mouse lens
12	785	85.7	522	5	RCAACRYST	X85205 R.catesbeia
13	759	82.9	1361	5	D86299	D86299 Cynops pyr
14	756	82.5	447	4	AJ617725	AJ617725 Elephas m
15	736	80.3	705	5	D88185	D88185 Xenopus lae
16	728.5	79.5	558	4	AJ617724	AJ617724 Ornithorh
17	715	78.1	477	5	AJ617726	AJ617726 Sphenodon
18	694	75.8	730	5	AY035778	AY035778 Danio rer
19	694	75.8	1430	5	BC083177	BC083177 Danio rer
20	677	73.9	448	5	TSU31938	TSU31938 Trachemys s
21	674	73.6	447	5	AJ617727	AJ617727 Lygodacty
22	663	72.4	448	5	EUDLAACRYS	L25850 Eudromia el
23	662	72.3	448	5	APRAAC	X96592 A.platyrrhin
24	653	71.3	448	5	CLRAAC	X96593 C.livia mRN
25	653	71.3	676	5	AY007972	AY007972 Clarias f
26	631	68.9	765	5	RTCRYA	X00716 Frog mRNA f
27	591	64.5	885	10	RNACRY	V01219 Partial seq
28	544	59.4	436	5	QLJ000940	AJ000940 Oryzias l
29	499	54.5	632	4	AF029793	AF029793 Bos tauru
30	493	53.8	548	4	OCRCYAB	X95383 O.cuniculus
31	489	53.4	528	9	BT006770	BT006770 Homo sapi
32	489	53.4	528	12	BT007909	BT007909 Synthetic
33	489	53.4	537	6	AX937702	AX937702 Sequence
34	489	53.4	537	6	AX937703	AX937703 Sequence
35	489	53.4	691	6	C0812349	C0812349 Sequence
36	489	53.4	691	6	AX330255	AX330255 Sequence
37	489	53.4	691	6	AX333032	AX333032 Sequence
38	489	53.4	691	6	BD079402	BD079402 Cancer-as
39	489	53.4	691	9	S45630	S45630 alpha B-cry
40	489	53.4	744	9	BC007008	BC007008 Homo sapi
41	489	53.4	856	6	AX888028	AX888028 Sequence
42	489	53.4	856	6	BD027638	BD027638 Sequence
43	489	53.4	893	6	AX899079	AX899079 Sequence
44	489	53.4	893	6	BD034612	BD034612 Sequence
45	489	53.4	911	6	AX899075	AX899075 Sequence
46	489	53.4	911	6	BD034608	BD034608 Sequence
47	489	53.4	927	6	AX899073	AX899073 Sequence
48	489	53.4	927	6	BD034606	BD034606 Sequence
49	489	53.4	942	6	AX888027	AX888027 Sequence
50	489	53.4	942	6	BD027637	BD027637 Sequence
51	489	53.4	1036	6	BD135185	BD135185 Human nuc
52	489	53.4	1036	6	AX017495	AX017495 Sequence
53	488	53.3	511	10	SEH272441	AJ272441 Spalax eh
54	488	53.3	528	10	SJU293658	AJ293658 Spalax ju
55	488	53.3	733	9	AB125159	AB125159 Macaca fa
56	485	52.9	661	9	AF007162	AF007162 Homo sapi
57	485	52.9	671	10	S74229	S74229 alpha B-cry
58	485	52.9	689	10	RRLNSABC	X60351 R.rattus mR
59	485	52.9	706	10	S77138	S77138 alpha B-cry
60	485	52.9	1247	6	AX401743	AX401743 Sequence
61	485	52.9	1247	10	RATCRYAB	M55534 Rat alpha-c
62	480	52.4	537	10	HAMSCARP	J03849 Hamster alp
63	480	52.4	687	10	RRHARTABC	X60352 R.rattus mR
64	480	52.4	704	10	S77142	S77142 alpha B-cry
65	479	52.3	522	5	RCABCRYST	X87114 R.catesbeia
66	476	52.0	666	10	MUSCRYABA	M63170 Mouse alpha
67	476	52.0	848	10	BC010768	BC010768 Mus muscu
68	476	52.0	966	5	BC082122	BC082122 Xenopus l
69	470	51.3	507	4	AJ617732	AJ617732 Elephas m
70	458.5	50.1	900	5	AF159089	AF159089 Danio rer
71	457.5	49.9	913	6	AX899074	AX899074 Sequence
72	457.5	49.9	913	6	BD034607	BD034607 Sequence
73	454.5	49.6	466	4	AJ617731	AJ617731 Macropus
74	444.5	48.5	466	4	AJ617730	AJ617730 Didelphis
75	441.5	48.2	577	5	AY007973	AY007973 Clarias b
76	439.5	48.0	689	5	DUKABC	L08078 Anas platyr
77	437.5	47.8	846	5	GGU26661	U26661 Gallus gall

78	437.5	47.8	1042	5	BX930014	Callus ga
79	436.5	47.7	1251	5	S53164	alpha B-cry
80	435.5	47.5	163962	9	BS000233	Pan trogl
81	435	47.5	148179	9	AP001631	Homo sapi
82	435	47.5	340000	9	AP001748	Homo sapi
83	430	46.9	888	5	BC075197	Xenopus l
84	427.5	46.7	501	4	AJ617728	Tachythr
85	407.5	44.5	502	4	AJ617729	Tachythr
86	406.5	44.4	826	6	AR415535	Sequence
87	406.5	44.4	826	6	AX972369	Sequence
88	406.5	44.4	826	6	BD111088	EST and e
89	404	44.1	249	10	RATACRYA	
90	401	43.8	249	10	RATACRYB	
91	400.5	43.7	6311	5	CHKCRYAA	
92	372.5	40.7	22895	2	AC084065	Mus muscu
93	370.5	40.4	1955	5	BC076518	Danio rer
94	367	40.1	419	9	AF026952	Homo sapi
95	366	40.0	16745	10	AC090881	Mus Muscu
96	349	38.1	1799	9	HSALPACRS	Y14789 H.sapiens a
97	343.5	37.5	1065	5	AY148412	Clarias b
98	343	37.4	906	10	MAACRY2	X02951 Hamster alp
99	343	37.4	166831	2	BX950201	Danio rer
100	342	37.3	2380	10	MCRYS	V00730 Mus musculu
101	342	37.3	2403	10	MUSCRYAA	J00375 Mouse lens
102	342	37.3	2580	10	MAACRY1	X02950 Hamster alp
103	341	37.2	1463	9	BC068046	BC068046 Homo sapi
104	340	37.1	1483	6	C0721193	Sequence
105	340	37.1	1457	9	AK056951	Homo sapi
106	340	37.1	1488	6	BD205158	Human nuc
107	340	37.1	1488	6	AX013767	Sequence
108	340	37.1	188810	2	AC121138	Mus muscu
109	337	36.8	165693	5	BX248514	Zebrafish
110	336	36.7	1310	10	D29960	Rattus norv
111	335	36.6	189	9	HUMCRYX1	M35629 Human alpha
112	330	36.0	635	6	AX888029	Sequence
113	330	36.0	695	6	BD027639	Sequence
114	321.5	35.1	4321	5	AFAACEPIG	Y11300 Astyanax fa
115	321	35.0	1316	10	SEHCRYAA2	M17249 Mole rat al
116	319.5	34.9	4339	5	AFAACCAVE	Y11301 Astyanax fa
117	317.5	34.7	765	5	RTCRYA	X00716 Frog mRNA f
118	316	34.5	5491	10	SEHCRYAA1	M17247 Mole rat al
119	308.5	33.7	719	6	CQ591247	Sequence
120	308.5	33.7	735	3	AY047516	AY047516 Drosophil
121	299	32.6	763	5	GGTAP	XY59541 Chicken mRN
122	296.5	32.4	771	3	AF315318	Bombyx mo
123	294.5	32.2	686	3	AF315319	Bombyx mo
124	294.5	32.2	764	3	AF315317	Bombyx mo
125	293	32.0	168	9	HUMAACA	L25781 Homo sapten
126	293	32.0	767	3	PIU94328	U94328 Plodia inte
127	292	31.9	535	10	MMU03562	U03562 Mus musculu
128	291	31.8	349	6	AX786923	Sequence
129	290	31.7	830	3	AF237691	Ciona int
130	289.5	31.6	615	9	CR407614	Homo sapi
131	289.5	31.6	618	9	CR536489	Homo sapi
132	289.5	31.6	724	9	BC014920	Homo sapi
133	289.5	31.6	764	6	CQ799993	Sequence
134	289.5	31.6	764	9	AB020027	Homo sapi
135	289.5	31.6	781	9	BC012768	Homo sapi
136	289.5	31.6	789	6	AC380755	Sequence
137	289.5	31.6	789	9	HSHSP27L	HS4079 Human mRNA
138	289.5	31.6	794	9	BC073768	Homo sapi
139	289.5	31.6	847	6	BD186273	STAY6 act
140	289.5	31.6	865	6	CQ730135	Sequence
141	289.5	31.6	865	9	HSU90906	Human clone
142	289.5	31.6	867	9	BC000510	Homo sapi
143	289.5	31.6	1231	6	AX411221	Sequence
144	289.5	31.6	1231	9	HSHSP28	Z23090 H.sapiens m
145	289.5	31.6	1380	6	AR217508	Sequence
146	289.5	31.6	1380	6	AX766880	Sequence
147	289	31.6	787	6	CQ858777	Sequence
148	289	31.6	787	6	AX401752	Sequence
149	289	31.6	787	10	RATHSP27A	M86389 Rattus norv
150	288.5	31.5	797	3	DMEL2131	X77635 Drosophila

ALIGNMENTS	
RESULT 1	
CR407691	519 bp mRNA linear PRI 10-MAY-2004
LOCUS	Homo sapiens full open reading frame cDNA clone RZPD0834E043D for gene CRYAA, crystallin, alpha A complete cds, without stopcodon.
DEFINITION	CR407691
ACCESSION	CR407691
VERSION	CR407691.1 GI:47115318
KEYWORDS	Full ORF shuttle clone, Gateway(TM), complete cds.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 519) Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.
TITLE	Cloning of human full open reading frames in Gateway(TM) system entry vector (pDONR201)
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 519) Ebert,L., Schick,M., Neubert,P., Schatten,R., Henze,S. and Korn,B.
AUTHORS	Direct Submission
TITLE	Submitted (07-MAY-2004) RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer Feid 580, D-69120 Heidelberg, Germany
COMMENT	RZPD: RZPD0834E043D, ORFNo 642 www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=RZPD0834E043D RZPDLIB: Human Full ORF Clones Gateway(TM) - RZPD (kan-resist.) RZPD LIB No. 834 www.rzpd.de/cgi-bin/products/showLib.pl.cgi/response?libNo=834 www.rzpd.de/products/orfclones/ Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Heubnerweg 6, D-14059 Berlin, Germany Tel: +49 30 32639 101 Fax: +49 30 32639 111 www.rzpd.de This clone is available from RZPD; contact RZPD (customer.service@rzpd.de) for further information. This CDS clone is a part of a collection of human full length expression clones generated by RZPD. This CDS has been cloned without stopcodon. The CDS has been inserted into pDONR201 via a BP Clonase(TM) reaction. Additional sequence has been added in front of the start codon (ATG): att. AAAAAA GGT GGC ACC CCT GGT CCA GGT (ATG) After the last codon additional sequence has been added: CCA GGC CCA GGC GGC G in front of the 3' att site (AC CCA GGT TTC TT). Compared to the reference sequence U66584 we did not find any amino acid exchanges. Clone distribution: http://www.rzpd.de/products/orfclones/.
FEATURES	Location/Qualifiers
source	1..519 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="RZPD0834E043D" /clone_lib="Human Full ORF Clones Gateway(TM) - RZPD" /lab_host="DH10B" /notes="Vector: pDONR201, Site_1: attP1; Site_2: attP2" 1..519 /gene="CRYAA" 1..519 /gene="CRYAA" /codon_start=1 /protein_id="CAG28619.1" /db_xref="GI:47115319" /translations="MDVITQHPMKRTLGPFPYPSRLFDQFSGELFYDLLPFLSSTI SPYRQSLPRTVLDSGIVSEVRDRDKFVFLDKVHFSPEDLTKVDQDFVEIHGKNE RQDDHGVISEFHRRYRLPSNVQVSALSCLSDAGMLTFCGPKIQTLGDATHAERAIP VSREKPTSPSS"
gene	
CDS	
ORIGIN	

Alignment Scores:

Pred. No.: 9,38e-94 Length: 519
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-657-740-1 (1-173) x CR407691 (1-519)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 1 ATGACGTCGACCATCAGACCCCTGGTCAAGCGCACCCCTGGGCCCTTCTACCCACG 60
 Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 61 CGGCTGTTTCGACCACTTTTCGGCGAGGGCCCTTTTGATGATGACTGCTGCCCTTCCTG 120
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 121 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCCGACCGCTGCTGACCTCCGCG 180
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallyHisPhe 80
 Db 181 ATCTCTGAGGTTCGATCCGACCGGACCAAGTTCGTCATCTTCTCGATGTGAAGCACTTC 240
 Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 241 TCCCGGAGGACCTCACCCTGAGGTGAGGACGACTTTTGGAGATCCACGGAAGCAC 300
 Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
 Db 301 AACGAGCCAGGACGACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTG 360
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 361 CCGTCCAAACGTGGACCACTCGGCCCTCTCTGCTCCCTGCTGCGGATGGATGCTGACC 420
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 421 TTCTGTGCCCCCAAGATCAGACTGGCTGGATGCCACCGACCGCGGAGGACCATCC 480
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 481 GTGTGCGGGAGGAGAGACCCACCTCGGCTCCCTCGTCC 519

RESULT 2

BC069528 614 bp mRNA linear PRI 30-JUN-2004
 LOCUS Homo sapiens crystallin, alpha A, mRNA (cdna clone MGC:96924
 DEFINITION IMAGE:7262133), complete cds.

ACCESSION BC069528.1 GI:46854598

VERSION BC069528.1

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 614) Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Strausberg, R.L., Feingold, B.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
 Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.F., Cabavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
 Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villaion, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Heiton, E., Kettelman, M., Madan, A., Rodriguez, S.,

Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 614)
 Strausberg, R.
 Direct Submission
 Submitted (29-APR-2004) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Baylor Human Genome Sequencing Center
 cDNA Library Preparation: Baylor Human Genome Sequencing Center
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (L1NL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleeged, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/L1NL at: <http://image.llnl.gov>
 Series: IRBR Plate: 1 Row: d Column: 7.

FEATURES

source
 1..614
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:96924 IMAGE:7262133"
 /tissue_type="PCR rescued clones"
 /clone_lib="NIH_MGC_244"
 /note="Vector: pPCR-Script Amp SK(+)"
 1..614
 /gene="CRYAA"
 /note="synonyms: CRYA1, HSPB4"
 /db_xref="LocusID:1409"
 /db_xref="MIM:123580"
 51..572
 /gene="CRYAA"
 /codon_start=1
 /product="crystallin, alpha A"
 /protein_id="AAH69528.1"
 /db_xref="GI:46854599"
 /db_xref="LocusID:1409"
 /db_xref="MIM:123580"

CDS

ORIGIN

Alignment Scores:
 Pred. No.: 1.16e-93 Length: 614
 Score: 916.00 Matches: 173
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-657-740-1 (1-173) x BC069528 (1-614)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20

```
|||||
51 ATGGATGTGACCATCAGACCCCTGGTTCAAGCGCACCCCTGGGCCCCTCTTACCCACG 110
QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 111 CGGCTGTTTCGACCACTTTTTCGGCGAGGCGCTTTTGTAGATATGACCTGCTGCTTCCTG 170
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 171 TCGTCCACATCAGCCCTTACTACCGCAGTCCCTCTTCCGACCGCTGCTGACTCCGCG 230
QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db 231 ATCTCTGAGTTCGATCCGCGGACCAAGTTCGTCATCTTCTCGATGTGAAGCACTTC 290
QY 81 SerProGluAspLeuThrValLysValGlnAspPhePheValGluIleHisGlyLysHis 100
Db 291 TCCCGGAGGACCTCACCGTGAAGGTGAGGACGACTTTGTGGAGATCCACGGAAGCAC 350
QY 101 ArgGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db 351 AACGAGCGCCAGGACGACACCGCTTACATTTCCCGTGAGTTCCACCGCGCTTACCGCTG 410
QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 411 CGGTCCACGTGACCACTCGGCCCTCTCTGCTCCCTGCTGCGATGGCATGCTGACC 470
QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 471 TTCTGTGGCCCCAGATCCAGACTGGCTGGATGCCACCCACCGCGGAGCGAGCCATCCCC 530
QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 531 GTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 569
```

```
RESULT 3
HSU66584 741 bp mRNA linear PRI 14-DEC-1996
LOCUS Human alphaA-crystallin (CRYAA) mRNA, complete cds.
```

```
DEFINITION U66584
```

```
ACCESSION U66584.1 GI:1732062
```

```
VERSION U66584.1
```

```
KEYWORDS Homo sapiens (human)
```

```
SOURCE Homo sapiens
```

```
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE 1 (bases 1 to 741)
```

```
AUTHORS Petrash,J.M., Mathur,S., Manoharan,M. and Andley,U.P.
```

```
TITLE Cloning and expression of human lens crystallins
```

```
JOURNAL Invest. Ophthalmol. Vis. Sci. 36, S882-S882 (1995)
```

```
REFERENCE 2 (bases 1 to 741)
```

```
AUTHORS Andley,U.P., Mathur,S., Griest,T.A. and Petrash,J.M.
```

```
TITLE Cloning, expression, and chaperone-like activity of human
```

```
alphaA-crystallin
```

```
J. Biol. Chem. 271 (50), 31973-31980 (1996)
```

```
JOURNAL J. Biol. Chem. 271 (50), 31973-31980 (1996)
```

```
MEDLINE 97112991
```

```
PUBMED 8943244
```

```
REFERENCE 3 (bases 1 to 741)
```

```
AUTHORS Petrash,J.M., Mathur,S., Wang,J.C., Griest,T.A. and Andley,U.P.
```

```
TITLE Direct Submission
```

```
JOURNAL Submitted (12-AUG-1996) Ophthalmology and Visual Sciences,
```

```
Washington University School of Medicine, 660 S. Euclid Ave., St.
```

```
Louis, MO 63110, USA
```

```
FEATURES
```

```
Location/Qualifiers
```

```
1..741
```

```
/organism="Homo sapiens"
```

```
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
```

```
/tissue_type="eye lens"
```

```
1..741
```

```
/gene="CRYAA"
```

```
1..522
```

```
/gene="CRYAA"
```

```
/function="structural protein"
/note="structurally related to sHSPs; chaperone-like
activity"
/codon_start=1
/product="alphaA-crystallin"
/protein_id="AAC50900.1"
/db_xref="GI:1732063"
/translation="MDVTIQHPFKRTLPFPSPRLPDPFEGELFEYDLLPFLSSTI
SPYRQSLPRTVLDSGISSEVRSDKVFILVDVKEFSPDLTKVQDDFVHKGHNE
RQDHGRTISREFHRRYRLPSPNVDSALSCSLSDAGMLTFCGPKIQTGLDATHAERAIP
VSREBKPTSPSS"
```

ORIGIN

```
Alignment Scores: 1.47e-93 Length: 741
Pred. No.: 916.00 Matches: 173
Score: 916.00 Conserv: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 9
```

```
US-10-657-740-1 (1-173) x HSU66584 (1-741)
```

```
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
```

```
Db 1 ATGGACGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGCGCCCTTCTTACCCACG 60
```

```
QY 21 ArgLeuPheAspGlnPhePheGlyGlyLeuPheGluTyrAspLeuProPheLeu 40
```

```
Db 61 CGGCTGTTTCGACCACTTTTTCGGCGAGGCGCTTTTGTAGATATGACCTCTGCCCTTCCTG 120
```

```
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
```

```
Db 121 TCGTCCACCATCAGCCCTTACTACCGCAGTCCCTCTTTCGGCACCGTCTGGACTCCGCG 180
```

```
QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
```

```
Db 181 ATCTCTGAGTTCGATCCGCGGACCAAGTTCGTCATCTTCTCGATGTGAAGCACTTC 240
```

```
QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
```

```
Db 241 TCCCGGAGGACCTCACCGTGAAGGTGAGGACGACACTTTTGTGGAGATCCACGGAAGCAC 300
```

```
QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
```

```
Db 301 AACGAGCGCCAGGACGACCAAGTTCATTTCCGCTGAGTTCACCGCGCGCTACCGCCTG 360
```

```
QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
```

```
Db 361 CGGTCCACGTGACCATCCAGTCCGCGCTCTTGTCTCCCTCTGCTGCGATGGCATGCTGACC 420
```

```
QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
```

```
Db 421 TTCTGTGGCCCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGGAGCGAGCCATCCCC 480
```

```
QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
```

```
Db 481 GTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 519
```

```
RESULT 4
```

```
LOCUS HSU05569
```

```
DEFINITION Human alphaA-crystallin (CRYA1) mRNA, complete cds.
```

```
ACCESSION U05569
```

```
VERSION U05569.1 GI:452477
```

```
KEYWORDS Homo sapiens (human)
```

```
SOURCE Homo sapiens
```

```
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE 1 (bases 1 to 1112)
```

```
AUTHORS Jaworski,C.J.
```

```
TITLE The human alphaA-crystallin gene
```

```
1112 bp mRNA linear PRI 25-APR-1996
```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Thesis (1992) LMDB, NEI, Molecular Structure and Function
2 (bases 1 to 1112)
Jaworski, C.J.
A reassessment of mammalian alpha A-crystallin sequences using DNA
sequencing: Implications for anthropoid affinities of tarsier
J. Mol. Evol. 41 (6), 901-908 (1995)
96139023
8587135
3 (bases 1 to 1112)
Wistow, G.J.
Direct Submission
Submitted (25-JAN-1994) Graeme J. Wistow, Molecular Structure and
Function, LMDB, NEI, NIH, Bethesda, MD 20892, USA
1. .1112
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/tissue_type="lens"
1. .1112
/gene="CRYA1"
68. .589
/gene="CRYA1"
/function="lens structural protein"
/note="This sequence is derived from genomic and cDNA
sequence"
/citation=[1]
/citation=[2]
/codon_start=1
/evidence=experimental
/product="alphaA-crystallin"
/protein_id="AA97523.1"
/db_xref="GI:452478"
/translation="MDVTIQHPKRTLGPFVPSRLPDQFFGEGLEFYDLLLPFLSSTI
SPYRQSLFRVLDSGISEVRSDKDFVILDKHFSPELITVKVQDDFVLIHGKNE
RQDDHGYISRFHRRYLRPLPSNVDSALSLSLADGMLTFCGFKITQGLDATHAERAI
VSREEKTPSAPSS"

Alignment Scores:
Pred. No.: 2.47e-93 Length: 1112
Score: 916.00 Matches: 173
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-657-740-1 (1-173) x HS005569 (1-1112)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 68 ATGACGTGACCATCCAGACCCCTGGTTCAGCGCACCTGGGGCCCTTCTACCCGAC 127
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 128 CGCGTGTTCGACCACTTTTCGGGAGGCGCTTTTGTAGTATGACCTCTGCCCTTCCTG 187
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 188 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCGACCGCTGCTGACCTCCGGC 247
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db 248 ATCTCTGAGGTTTCGATCCGACCGGACAAAGTTCGTCACTTCTCGATGTGAGCACTTC 307
Qy 81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
Db 308 TCCCGGAGGACCTCACCGTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 367
Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db 368 AACGAGCGCCAGGACGACCGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCGCTG 427

ORIGIN

Alignment Scores:
Pred. No.: 2.47e-93 Length: 1114
Score: 916.00 Matches: 173
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-657-740-1 (1-173) x CQ731849 (1-1114)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 70 ATGACGTGACCATCCAGACCCCTGGTTCAGCGCACCTGGGGCCCTTCTACCCGAC 129
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 130 CGCGTGTTCGACCACTTTTCGGGAGGCGCTTTTGTAGTATGACCTCTGCCCTTCCTG 189
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 190 TCGTCCACCATCAGCCCTTACTACCGCCAGTCCCTCTTCGACCGCTGCTGACCTCCGGC 249
Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
Db 250 ATCTCTGAGGTTTCGATCCGACCGGACAAAGTTCGTCACTTCTCGATGTGAGCACTTC 309
Qy 81 SerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHis 100
Db 310 TCCCGGAGGACCTCACCGTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 369
Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db 370 AACGAGCGCCAGGACGACCGGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCGCTG 429
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140

Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 428 CCGTCCACGCTGGACCATCGGCCCTCTCTTGTCTCCCTGTGCGATGGATGCTGACC 487
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 488 TTCTGTGCCCCCAAGATCCAGACTGGCTGTGATGCCACCCACGCGGAGGACCATCCCC 547
Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 548 GTGTGCGGGAGGAGAGGCCACCTCGGCTCCCTCGTCC 586

RESULT 5
CQ731849
LOCUS
DEFINITION
Sequence 17783 from Patent WO20068579.
ACCESSION
CQ731849
VERSION
CQ731849.1 GI:42309516
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL
Patent: WO 02068579-A 17783 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source
1. .1114
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

Db 430 CCGTCCACGTTGGAGCAGCTCGGCCCTCTCTTGGCTCCCTGTCTGCGGATGCGTGCACC 489
 Qy 141 PheCysGlyProLysIleGlnThrClyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 490 TTCTGTGGCCCCAAGATCCACACTGGCTGGATGCCACCCACGCGGAGGAGCCATCCCC 549
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 550 GTGTGCGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 588

RESULT 6

OCCRYAA
 LOCUS O.cuniculus mRNA for alpha-A-crystallin. 543 bp mRNA linear MAM 23-OCT-1996
 ACCESSION X95382

VERSION X95382.1 GI:1177576
 KEYWORDS alpha-A-crystallin; cryaA gene.
 SOURCE Oryctolagus cuniculus (rabbit)

ORGANISM

Oryctolagus cuniculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

REFERENCE

AUTHORS Krausz, E., Augusteyn, R.C., Quinlan, R.A., Reddan, J.R., Russell, P.,
 Sax, C.M. and Graw, J.

TITLE Expression of Crystallins, Pax6, Filensin, CP49, MIP, and MP20 in
 lens-derived cell lines

JOURNAL Invest. Ophthalmol. Vis. Sci. 37 (10), 2120-2128 (1996)

MEDLINE 96409169

PUBMED 8814151

REFERENCE 2 (bases 1 to 543)

AUTHORS

Graw, J.

TITLE Direct Submission

JOURNAL Submitted (29-JAN-1996) J. Graw, Institute of Mammalian Genetics,
 GSF-Research Center Neuherberg, Ingolstaedter Landstrasse 1,
 Oberschleissheim, D-85764, FRG

FEATURES

source

1..543 /organism="Oryctolagus cuniculus"

/mol_type="mRNA"

/sub_species="New Zealand white"

/db_xref="taxon:9986"

/tissue type="lens"

gene

1..543 /gene="cryaa"

/gene="cryaa"

22..543 /gene="cryaa"

/codon_start=1

/product="alpha-A-crystallin"

/protein_id="CAA64668.1"

/db_xref="GI:1177577"

/db_xref="GOA:P02493"

/db_xref="UniProt/Swiss-Prot:P02493"

/translation="MDVTIQHPWFKRLGPPYPSRLFDQFFGGLFEYDLLPFLSSTI
 SPYVRSGLFRTVLDSDGISEVRSDRDKFVFLDVKHPSPEDLTAVKQDFVFIHGKHN
 RQDDHGYSREFHRRYRLPSNVDSALSCSLSDGMLTFSGPKVQSGLDAGHSRAIP
 VSREKPSVPSS"

ORIGIN

Alignment Scores:

Pred. No.: 1..47e-89 Length: 543

Score: 879.00 Matches: 165

Percent Similarity: 98.27% Conservative: 5

Best Local Similarity: 95.38% Mismatches: 3

Query Match: 95.96% Indels: 0

DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x OCCRYAA (1-543)

Qy 1 MetAspValThrIleGlnHisProThrPheLysArgThrLeuGlyProPheTyProSer 20

Db 22 ATGGACGTCCACATCCAGCACCCCTGGTTCAGGCGCACCTGGGGGCCCTTCTACCCGAGC 81

Qy 21 ArgLeuPheAspGlnPhePheGlyGluLeuPheGluTyAspLeuLeuProPheLeu 40
 Db 82 CGGCTGTTCCGACGAGTTCTTTCGTTGAGGGCTCTTCGAGTATGACCTGCTGCCCTTCCTG 141
 Qy 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 142 TCGTCCACCATCAGCCCTACTACCGCCAGTCCCTCTTCGCGACCGTGTGGACTCCGGC 201
 Qy 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 Db 202 ATCTCTGAGTGGCTCCGACCGGACAGTTCTGTCATCTTCTCGGACGTGAGCACTTC 261
 Qy 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 Db 262 TCGCCCGGAGACCTCACCGTGAAGTGCAGGAGACTTCGTGGAGATCCACGGCAACAC 321
 Qy 101 AsnGluArgGlnAspAspHisGlyTyIleSerArgGluPheHisArgArgTyArgLeu 120
 Db 322 AACGAGACAGGATGACCGGCTACATTTCCCGCGAGTTCCACCGCGCTACCGCTG 381
 Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 Db 382 CATTCCAACTGGACCACTGCGGCTGCTGCTGCTCTCCGCGGACGGCATGCTACCC 441
 Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
 Db 442 TTCTCCGCGCCCAAGGTGCAGTCCGGCTGAGCGCTGGCCACAGCGAGGAGCATCCCC 501
 Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 502 GTGTCCCGGAGGAGAGCCAGCTCGTGGCTCCCTCGTCC 540

RESULT 7

RNU47922
 LOCUS Rattus norvegicus alpha A-crystallin mRNA, complete cds. 1056 bp mRNA linear ROD 02-APR-1996
 DEFINITION Rattus norvegicus alpha A-crystallin mRNA, complete cds.
 ACCESSION U47922
 VERSION U47922.1 GI:1245161

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE 1 (bases 1 to 1056)

AUTHORS Bhat,S.P., Nandy,P., Srinivasan,A., Cheng,D. and Sitay,A.

TITLE Direct Submission

JOURNAL Submitted (31-JAN-1996) Suraj P. Bhat, JSEI, UCLA School Medicine,

100-Stein Plaza, Rm. BH623, Los Angeles, CA 90095-7008, USA

FEATURES

Location/Qualifiers

1..1056

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

13..534

/note="similar to small heat shock proteins"

/codon_start=1

/product="alpha A-crystallin"

/protein_id="AAA93367.1"

/db_xref="GI:1245162"

/translation="MDVTIQHPWFKRALGPPYPSRLFDQFFGGLFEYDLLPFLSSTI
 SPYVRSGLFRTVLDSDGISEVRSDRDKFVFLDVKHPSPEDLTAVKQDFVFIHGKHN
 RQDDHGYSREFHRRYRLPSNVDSALSCSLSDGMLTFSGPKVQSGLDAGHSRAIP
 VSREKPSVPSS"

535..1050

3'UTR.

ORIGIN

Alignment Scores:

Pred. No.: 2.73e-88 Length: 1056

Score: 871.00 Matches: 164

Percent Similarity: 97.69% Conservative: 5

Best Local Similarity: 94.80% Mismatches: 4

Query Match: 95.09% Indels: 0
 DB: 10 Gaps: 0

US-10-657-740-1 (1-173) x RNU47922 (1-1056)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 DB 13 ATGGACGTACCATCAGCACCCCTGGTTCAAGCGCGCCCTGGGCCCTTCTACCCAGC 72
 QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 DB 73 CGACTGTTTCGACCACTTCTTCGGCGAGGGCCCTTTTGAATACGACTGCTGCCCTTCTCTG 132
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 DB 133 TCTTCCACCATCAGCCCTCTACTACCGCCAGTCTCTCTTCGCGACAGTGTGGACTCCGGC 192
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 DB 193 ATCTCTGAGGTCGATCTGACCGGACCAAGTTTGTTCATCTTCTGGATGTGAAGCACTTC 252
 QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 DB 253 TCTCTCTGAGGACCTCAGCGTGAAGTACTGGAGATTTCGTGGAGATCATGGCAACAC 312
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 DB 313 AACGAGAGCAGGATGACCATGGCTACATTTCCTGGATTTCACCGTCGCTACCGCTCTG 372
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 DB 373 CTTTCCAATGTGGACCACTCGCCCTCTCTGCTCTCTGCTGCGATGCGATGCTGACC 432
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaLeuPro 160
 DB 433 TTCTCTGCGCCCAAGTCCAGTCTGGCTGGATGCTGGCCACGAGGAGGCGCACTTCCC 492
 QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 DB 493 GTGTCAAGGAGGAGAGAGCCAGCTCGCACCTCGTCC 531

RESULT 8
 BOVCRYA
 LOCUS 793 bp mRNA linear MAM 26-APR-1993
 DEFINITION Bovine alpha-A-crystallin gene, complete cds.
 ACCESSION M26142
 VERSION M26142.1 GI:162909
 KEYWORDS crystallin.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.
 REFERENCE 1 (bases 1 to 793)
 AUTHORS Hay, R.B. and Petrash, J.M.
 TITLE Nucleotide sequence of a bovine lens alpha A-crystallin cDNA
 JOURNAL Biochem. Biophys. Res. Commun. 148 (1), 31-37 (1987)
 MEDLINE 88049675
 PUBMED 3675580
 COMMENT Original source text: Bovine lens, cDNA to mRNA, clone pBL-alpha-A2-1.
 FEATURES
 source
 1. 793
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 31. 552
 /note="alpha-A-crystallin"
 /codon_start=1
 /protein_id="AAA30471.1"
 /db_xref="GI:162910"
 /translation="MDIAIQHPWFKRTIGPYPRLFDQFFEGGLFEYDLAPFLSSTI
 SPYVRSQSLFRTVDSGISEVSRDRDKFVIFLDVKHFSPEDLTVKVKQEDFVEIHGKHE

ORIGIN 29 bp upstream of NcoI site.

Alignment Scores:
 Pred. No.: 3,198-88 Length: 793
 Score: 869.00 Matches: 163
 Percent Similarity: 97.69% Conservative: 6
 Best Local Similarity: 94.22% Mismatches: 4
 Query Match: 94.87% Indels: 0
 DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x BOVCRYA (1-793)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
 DB 31 ATGGATATCGCATTCAGCACCCCTGGTTCAACGACCCCTGGGCCCTTCTACCCAGC 90
 QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 DB 91 CGGCTGTTTCGACCACTTCTTCGGCGAGGGCCCTCTTCGAGTACGACCTGCTGCCCTTCTCTG 150
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 DB 151 TCTTCCACCATCAGCCCTCTACTACCGCCAGTCTCTTTCGCGACCGTGTGACTCCGGC 210
 QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
 DB 211 ATCTCTGAGTCCGATCCGACCGGACCAAGTTTGTTCATCTTCTGGATGTGAAGCACTTC 270
 QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
 DB 271 TCTCCGAGGACCTGACGGTGAAGTGCAGGAGGACTTCGTGGAGATCCACGGCAAGCAC 330
 QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
 DB 331 AACGAGCGGAGGATGACCATGGCTACATCTCCGCGAGTTCACCGCGCTACCGCTG 390
 QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
 DB 391 CTTTCCAAGTGGACCACTCGCATCTCTCTGCTCTCTGCTGCTGATGGCATGCTGACC 450
 QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaLeuPro 160
 DB 451 TTCTCTGCGCCCAAGATCCATCTCGGTGAGCGCGCCACAGCGAGGCGGCGCATCCGCC 510
 QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
 DB 511 GTGTCCCGGAGGAGAGCCAGCTCTGCGCCCTCGTCC 549

RESULT 9
 MMU310308
 LOCUS 796 bp mRNA linear ROD 03-APR-2001
 DEFINITION Mus musculus mRNA for alpha-A-crystallin (Cryaa gene).
 ACCESSION AJ310308
 VERSION AJ310308.1 GI:13548627
 KEYWORDS alpha-A-crystallin; cryaa gene.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 Graw, J.
 TITLE Characterization of a new, dominant V124E mutation in the mouse
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 796)
 Graw, J.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-2001) Graw J., Institute of Mammalian Genetics,
 GSF-National Research Center for Environment and Health,
 Ingolstaedter Landstr. 1, Neuherberg, Germany, D-85764, GERMANY
 FEATURES
 Location/Qualifiers

```

source
1..796
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C3H/El"
/db_xref="taxon:10090"
/tissue_type="lens"
1..796
/gene="Cryaa"
52..642
/gene="Cryaa"
/function="lens structural protein; chaperone"
/feature="alternative splicing"
/codon_start=1
/product="alpha-A-crystallin"
/protein_id="CAC35974.1"
/db_xref="GI:13548628"
SPYRQSLFRTVLDSGISELTHMFMVHQHAGNPKNNPVKVRSDRDKFVIFLDVKH
FSPEDLTVKVLEDFVEIHGKNERQDDHGYISREFHRRYRLPSNVDSQALSCLSDAG
MLTFSGPKVQSLDAGHSERAI PVSREKPSAPSS"
241..309
/exon
/gene="Cryaa"
/feature="alternative splicing"

ORIGIN
Alignment Scores:
Pred. No.: 5,04e-86 Length: 796
Score: 849.50 Matches: 164
Percent Similarity: 86.22% Conservative: 5
Best Local Similarity: 83.67% Mismatches: 4
Query Match: 92.74% Indels: 23
DB: 10 Gaps: 1

US-10-657-740-1 (1-173) x MMU310308 (1-796)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 52 ATGGAGCGTCACCATTCAGCATCTTGGTTCAAGCGTGCCTGGGGCCCTTCTACCCGAGC 111
Qy 21 ArgLeuPheAspGlnPhePheGlyGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db 112 CGACTGTTTCGACCACTTCTTCGGCGAGGCGCTTTTGTAGTACGACCTGCTGCCCTTCCTG 171
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 172 TCTTCCACATCAGCCCTTACTACCGCCAGTCCTTCCGACCTGCTGCTGCTGCTGCGGC 231
Qy 61 IleSerGlu----- 63
Db 232 ATCTCTGAGCTCATGACCATATGCGTTTGTAAATGCACCAACACCATGCTGGAACCCC 291
Qy 64 -----ValArgSerAspArgAspLysPheValIlePheLeuAspVal 77
Db 292 AAGAACAACCCCGTCAGGTCGATCTCAGCGGACAGATTTGTTCATCTTCTTGACGCTG 351
Qy 78 LysHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGluIleHis 97
Db 352 AAGCACTTCTCTCTCGAGACCTCACCGTGAAGGTACTGGAGGATTTTGTGGAGATTCCAC 411
Qy 98 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 117
Db 412 GGCAACAACACAGAGAGGAGGATGACCATGGCTACATTTCCCGTGATTTCCAGCTGCG 471
Qy 118 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
Db 472 TACCGTCTGCGCTTCCAAATGTGAGACAGTCCGCGCTCTCTCGCTCCCTGCTGCGGATGCG 531
Qy 138 MetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
Db 532 ATGCTGACCTTCTCTGGCCCCAAGTCCAGTCCGGTTTGGATGCTGGCCACAGGAGAGG 591
Qy 158 AlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173

```

```

Db 592 GCCATTCTGTGTACGGAGGAGAGAAACCCAGCTCTGCACCCCTCGTCC 639
RESULT 10
RNU47921
LOCUS RNU47921 1271 bp mRNA linear ROD 02-APR-1996
DEFINITION Rattus norvegicus alpha A (insert)-crystallin mRNA, complete cds.
ACCESSION U47921
VERSION U47921.1 GI:1245159
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 1271)
AUTHORS Bhat,S.P., Nandy,P., Srinivasan,A., Cheng,D. and Sitay,A.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-1996) Suraj P. Bhat, JSEI, UCLA School Medicine,
100-Stein Plaza, Rm. BH623, Los Angeles, CA 90095-7008, USA
FEATURES
Location/Qualifiers
source
1..1271
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="lens"
159..749
CDS
/feature="alpha A (insert)-crystallin is similar to alpha
A-crystallin, the predominant protein of the ocular lens.
These two proteins are identical in sequence except for
the presence of a 23 amino acid peptide in alpha
A(insert)-crystallin located between the 63rd and the 64th
residue of alpha A-crystallin; alpha A(insert)-crystallin
represents about 10-20% of total alpha crystallin in the
lens"
/codon_start=1
/product="alpha A (insert)-crystallin"
/protein_id="AAA93366.1"
/db_xref="GI:1245160"
/translation="MDVTIOHPFKRALGPFYPSRLFDQFEGELFYDILLPFLSSSTI
SPYRQSLFRTVLDSGISELTHMFMVHQHAGNPKNNPVKVRSDRDKFVIFLDVKH
FSPEDLTVKVLEDFVEIHGKNERQDDHGYISREFHRRYRLPSNVDSQALSCLSDAG
MLTFSGPKVQSLDAGHSERAI PVSREKPSAPSS"
348..416
misc_feature
/feature="alternatively spliced region; alpha A
(insert)-crystallin mRNA is produced by alternative
splicing from the alpha A-crystallin gene; insert peptide
sequence: LMTHMFMVHQHAGNPKNNPGK"
750..1271
3'UTR
ORIGIN
Alignment Scores:
Pred. No.: 9.13e-86 Length: 1271
Score: 849.50 Matches: 164
Percent Similarity: 86.22% Conservative: 5
Best Local Similarity: 83.67% Mismatches: 4
Query Match: 92.74% Indels: 23
DB: 10 Gaps: 1

US-10-657-740-1 (1-173) x RNU47921 (1-1271)
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 159 ATGGAGCTCACCATTCAGCACCTTGGTTCAAGCGGCGCCCTTCTACCCGAGC 218
Qy 21 ArgLeuPheAspGlnPhePheGlyGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db 219 CGACTGTTTCGACCACTTCTTCGGCGAGGCGCTTTTGAATACGACCTGCTGCCCTTCCTG 278
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 279 TCTTCCACCATCAGCCCTTACTACCGCCAGTCTCTCTTCGCGACAGTGTGGACTCCGCG 338

```

Qy 61 IleSerGlu----- 63
 Db 339 ATCTCTGAGCTCATGACCCATATGTGTTTGAATGACCAACACATGCTGGAACCCC 398
 Qy 64 -----ValArgSerAspArgAspLysPheValIlePheLeuAspVal 77
 Db 399 AAGAACACCCCGGCAAGTCCGATCTGACCGGGACAAGTTTGTCTCTTGTGATGTG 458
 Qy 78 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 97
 Db 459 AAGCACTTCTCTCTGAGGACCTACCGTGAAGGTACTGGAAGATTTCGTGGAGATCCAT 518
 Qy 98 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 117
 Db 519 GGCACACACACAGAGAGGAGGATGACCATGCTACATTTCOCGTGAATTCACCGTCGC 578
 Qy 118 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
 Db 579 TACCGTCTGCGCTTCCAAATGTGACAGTCCGCGCCCTCTCTGCTCTGCTGCGATGCG 638
 Qy 138 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
 Db 639 ATGTGACTTCTCTGCGCCCAAGTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
 Qy 158 AlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 699 GCCATTCCGTTGTCACGGGAGGAGAGCCAGCTCGGCACCCCTCGTCC 746

RESULT 11
 MUSCRYAAB 1025 bp mRNA linear ROD 27-APR-1993
 LOCUS Mouse lens alpha-A-crystallin mRNA, 3' end.
 DEFINITION J00376
 ACCESSION J00376.1 GI:192760
 VERSION alpha-crystallin; crystallin.
 KEYWORDS Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1025)
 King, C.R., Shinohara, T. and Piatigorsky, J.
 alpha-A-crystallin messenger RNA of the mouse lens: more noncoding
 than coding sequences
 Science 215 (4535), 985-987 (1982)

JOURNAL 7156978
 MEDLINE
 PUBMED
 COMMENT
 Original source text: Mouse lens, cDNA to mRNA.
 The crystallins contribute 90% of the soluble lens protein and are highly conserved during evolution. There are four immunologically separate classes of crystallins called alpha-beta-, gamma-, and delta-crystallin. The mRNA of the alpha-A-crystallin from mouse, like that from cows and rats, is nearly three times the size required to code for the polypeptide. Nucleotides 1-490 of the mouse alpha-A-crystallin mRNA predict a protein sequence identical to amino acids 10-173 of rat alpha-A-crystallin.

FEATURES
 Source
 1..1025
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 <1..1025
 /product="CRY-alpha-A mRNA"
 <1..492
 /note="alpha-A-crystallin"
 /codon_start=1
 /protein_id="AAA37471.1"
 /db_xref="GI:387134"
 /translation="KRALGPFVPSRLPDPFGEGFVFDLLPFLSSTISPYVROSRLR
 TVLDSGISEVSDRDKFVFLDVKHFSPEDLTKVLEDFVIEHKGHNERQDDHYISR
 EFHRRYRPNVDQSALSCSADGMLTFSGPKVQSGLDAGHSRAIPVSRKPPSSA
 PSS"
 ORIGIN 70 bp upstream of Real site.

Alignment Scores:
 Pred. No.: 1.97e-81 Length: 1025
 Score: 810.00 Matches: 154
 Percent Similarity: 97.55% Conservative: 5
 Best Local Similarity: 94.48% Mismatches: 4
 Query Match: 88.43% Indels: 0
 Db: 10 Gaps: 0

US-10-657-740-1 (1-173) x MUSCRYAAB (1-1025)

Qy 11 LysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhePheGlyGluGly 30
 Db 1 AAGCGTCCCTGGGGCCCTTCTACCCAGCCAGCTGTTGACCACTTCTTCGGCAGGGC 60
 Qy 31 LeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIleSerProTyrTyrArgGln 50
 Db 61 CTTTTGAGTACGACCTGCTGCTCTCTCTCTCCACCATCAGCCCTACTACCGCCAG 120
 Qy 51 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 70
 Db 121 TCCCTCTTCCGCACTGCTGCACTCGGGCATCTCTGAGGTCCGTTCCGACCGGACAAG 180
 Qy 71 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 90
 Db 181 TTTGTCTCTTCTTGGAGCTGAAGCACTTCTCTCTGAGGACCTCACCCTGAAAGTACTG 240
 Qy 91 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 110
 Db 241 GAAGATTTTGGAGATTCAGGCACACACACAGAGGAGGATGACCATGGCTACATA 300
 Qy 111 SerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 130
 Db 301 TCCCGTGAATTCACCGTGGTACCGTCTGCTCTCCAAATGTGGACCACTCGCCCTCTCC 360
 Qy 131 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 150
 Db 361 TGCTCCCTGCTGCTGATGGCATGTCACCTTCTCTGCCCCCAAGTCCAGTCCGGTTG 420
 Qy 151 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 170
 Db 421 GATCTGCGCACAGGAGAGGCCATTCCTGTCTACGGAGGAGAAACCCAGCTCTGCA 480
 Qy 171 ProSerSer 173
 Db 481 CCCTCGTCC 489

RESULT 12
 RCAACRYST 522 bp mRNA linear VRT 18-SEP-1995
 LOCUS R.catesbeiana mRNA for alpha-crystallin alpha A subunit.
 DEFINITION R.catesbeiana
 ACCESSION X85205
 VERSION X85205.1 GI:732915
 KEYWORDS aa-crystallin.
 SOURCE Rana catesbeiana (bullfrog)
 ORGANISM Rana catesbeiana
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana;
 Aquarana.

REFERENCE 1
 Lu, S.F., Pan, F.M. and Chiou, S.H.
 Sequence analysis of frog alpha-crystallin cDNA and its deduced
 primary structure: comparison of alpha A subunit chains among
 different vertebrate species
 Biochem. Biophys. Res. Commun. 210 (3), 974-981 (1995)
 JOURNAL 95283564
 MEDLINE 7763271
 PUBMED
 REFERENCE 2 (bases 1 to 522)
 Chiou, S.H.
 Direct Submission
 Submitted (10-MAR-1995) S.H. Chiou, Institute of Biological
 Chemistry, Academia Sinica, Institute of Biochemical Sciences,
 National Taiwan University, P.O. Box 23-106, Taipei 10099, TAIWAN
 Sequence overlapping with that under the acc#X00716.
 COMMENT

```
FEATURES
  source
    Location/Qualifiers
      1..522
        /organism="Rana catesbeiana"
        /mol_type="mRNA"
        /db_xref="taxon:8400"
      1..522
        /codon_start=1
        /product="alpha-crystallin alpha A subunit"
        /protein_id="CAA59471.1"
        /db_xref="GI:732916"
        /db_xref="GOA:Q91311"
        /db_xref="UniProt/Swiss-Prot:Q91311"
        /translation="MDTAIQHPFKRALGPPYNNRFDQVFGEGMFDYDLFPFLSSVT
        SPYVRSUFRGMSGISEVSDRDRFTINLDVKHFSDDLTVKILDDFVHIGKHS
        RODHGYVSRFHRRLPASVDQSSITCSLSTGMLTFSGTKMQSSLDSSHGRPIP
        VSREKPTSAQSS"
      CDS
        1..522
          /product="new alpha A-crystallin"
          /protein_id="BAB85811.1"
          /db_xref="GI:19168452"
          /translation="MDTIQHPFKRALGPPYNNRFDQVFGDGLFYELFPFLSSVT
          SPYVRSUFRNLDGISEVSGRDKFOIYLDVKHFSPEDLVKILDDYVHIGKHS
          RODHGYVSRFHRRLPASVDQSSITCSLSTGMLTFSGTKMQSSLDSSHGRPIP
          VSREKPSAQSS"
      ORIGIN
        Alignment Scores:
          Pred. No.: 1.58e-75 Length: 1361
          Score: 759.00 Matches: 135
          Percent Similarity: 90.75% Conservative: 22
          Best Local Similarity: 78.03% Mismatches: 16
          Query Match: 82.86% Indels: 0
          DB: 5 Gaps: 0
        US-10-657-740-1 (1-173) x RCAAACRYST (1-522)
        QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
        DB 1 ATGGACATCGGCATCCAGCACCCCTGGTTCAAGCGTGCTCTGGGGCCCTTCTACCCCAAC 60
        QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyAspLeuLeuProPheLeu 40
        DB 61 CGCCTCTTCGATCAGTCTCGAGAGAAATGTTTGACTAGCCTCTTTCCTTTCTCG 120
        QY 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGly 60
        DB 121 TCCTCCACCGTCAGTCCCTACTACAGGCACAGCCCTCTCCGAGGATTCATGGACTCCGGC 180
        QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
        DB 181 ATTTCTGAGTGCGTTCAGATCGCATCGTTTACCATTAACTTGATGTGAAGCATTTTC 240
        QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
        DB 241 TCACCCGATGATCTGACTGTGAAGATCTTGATGACTTTTGGAATCCATGGAAAGCAC 300
        QY 101 AsnGluArgGlnAspHisGlyTyIleSerArgGluPheHisArgArgTyArgLeu 120
        DB 301 AGCGAGAGCGAGGATGACCATGGCTACATTTCCCGGAGTTCACCGCGCGCTACCGCTC 360
        QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
        DB 361 CCTTCCAACTGGACCAATCTCTATCAGTGTCTCCTCTGCTGATGGCATTTTGACC 420
        QY 141 PheCysGlyProLysIleGlnThrChlYLeuAspAlaThrHisAlaGluArgAlaIlePro 160
        DB 421 TTCACGCGACCAAAATATGTCAGGCGCTGGACTCCAGCCACAGCGAGAGCCTATCCCT 480
        QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
        DB 481 GTGTCCAGAGAGAAAGCCCACTCTGCCCTCTCTCC 519
        RESULT 13
        LOCUS D86299
        DEFINITION Cynops pyrrhogaster mRNA for new alpha A-crystallin, complete cds.
        ACCESSION D86299
        VERSION D86299.1 GI:19168451
        KEYWORDS Cynops pyrrhogaster (Japanese firebelly newt)
        SOURCE Cynops pyrrhogaster
        ORGANISM Cynops pyrrhogaster
```

```
REFERENCE
  AUTHORS
  TITLE
    JOURNAL
    MEDLINE
    PUBMED
  REFERENCE
    Mizuno, N., Agata, K., Sawada, K., Mochii, M. and Eguchi, G.
    Expression of crystallin genes in embryonic and regenerating newt
    lenses
    Dev. Growth Differ. 44 (3), 251-256 (2002)
    Mizuno, N.
    Direct Submission
    Submitted (27-JUN-1996) Nobuhiko Mizuno, Osaka University,
    Institute for Molecular and Cellular Biology, Kondoh
    Differentiation Signaling Project ERATO; 1-3 Yamadaoka, Suitashi,
    Osaka 565-0871, Japan (E-mail: nmizuno@imcb.osaka-u.ac.jp,
    Tel: 06-6877-7964, Fax: 06-6877-1738)
  FEATURES
    source
      1..1361
        /organism="Cynops pyrrhogaster"
        /mol_type="mRNA"
        /db_xref="taxon:8330"
      142..663
        /codon_start=1
        /product="new alpha A-crystallin"
        /protein_id="BAB85811.1"
        /db_xref="GI:19168452"
        /translation="MDTIQHPFKRALGPPYNNRFDQVFGDGLFYELFPFLSSVT
        SPYVRSUFRNLDGISEVSGRDKFOIYLDVKHFSPEDLVKILDDYVHIGKHS
        RODHGYVSRFHRRLPASVDQSSITCSLSTGMLTFSGTKMQSSLDSSHGRPIP
        VSREKPSAQSS"
      ORIGIN
        Alignment Scores:
          Pred. No.: 1.58e-75 Length: 1361
          Score: 759.00 Matches: 135
          Percent Similarity: 90.75% Conservative: 22
          Best Local Similarity: 78.03% Mismatches: 16
          Query Match: 82.86% Indels: 0
          DB: 5 Gaps: 0
        US-10-657-740-1 (1-173) x D86299 (1-1361)
        QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
        DB 142 ATGGACATCGGCATCCAGCACCCCTGGTTCAAGCGGCTTCCGAGCCTTTACCTTGT 201
        QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyAspLeuLeuProPheLeu 40
        DB 202 CGCCTCTTCGACCAAGTCTTCGGGATGGCTCTTCGACTACGAGCTCTTCCCTTCTTA 261
        QY 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGly 60
        DB 262 TCCTCCACCGTCAGTCCCTACTACCGCAGTCCATGTTCCGCAACTACTCGACTCTG 321
        QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
        DB 322 ATATCCGAGGTGAGTCTGCGCGCAGCAAGTTCCAGATTTCACCTGGAGCTCAAGCACTTC 381
        QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
        DB 382 TCACCGAGAGACCTAAGTCAAGATCCTCGACGATTACGTGGAGATCCATGGAAGCAC 441
        QY 101 AsnGluArgGlnAspHisGlyTyIleSerArgGluPheHisArgArgTyArgLeu 120
        DB 442 AGCGACCGCGAGGATGATCAGCGTATGCTCTCGGAGAGTTCACCGCGCTATCGCCTG 501
        QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
        DB 502 CCGCGCGCGTGGACCACTCTCCATGCTCCCTGTCTGCTGCTGCTGCTGCTGCTGCTG 561
        QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
```

```
FEATURES
  source
    Location/Qualifiers
      1..522
        /organism="Rana catesbeiana"
        /mol_type="mRNA"
        /db_xref="taxon:8400"
      1..522
        /codon_start=1
        /product="alpha-crystallin alpha A subunit"
        /protein_id="CAA59471.1"
        /db_xref="GI:732916"
        /db_xref="GOA:Q91311"
        /db_xref="UniProt/Swiss-Prot:Q91311"
        /translation="MDTAIQHPFKRALGPPYNNRFDQVFGEGMFDYDLFPFLSSVT
        SPYVRSUFRGMSGISEVSDRDRFTINLDVKHFSDDLTVKILDDFVHIGKHS
        RODHGYVSRFHRRLPASVDQSSITCSLSTGMLTFSGTKMQSSLDSSHGRPIP
        VSREKPTSAQSS"
      CDS
        1..522
          /product="new alpha A-crystallin"
          /protein_id="BAB85811.1"
          /db_xref="GI:19168452"
          /translation="MDTIQHPFKRALGPPYNNRFDQVFGDGLFYELFPFLSSVT
          SPYVRSUFRNLDGISEVSGRDKFOIYLDVKHFSPEDLVKILDDYVHIGKHS
          RODHGYVSRFHRRLPASVDQSSITCSLSTGMLTFSGTKMQSSLDSSHGRPIP
          VSREKPSAQSS"
      ORIGIN
        Alignment Scores:
          Pred. No.: 5.48e-79 Length: 522
          Score: 785.00 Matches: 141
          Percent Similarity: 92.49% Conservative: 19
          Best Local Similarity: 81.50% Mismatches: 13
          Query Match: 85.70% Indels: 0
          DB: 5 Gaps: 0
        US-10-657-740-1 (1-173) x RCAAACRYST (1-522)
        QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
        DB 1 ATGGACATCGGCATCCAGCACCCCTGGTTCAAGCGTGCTCTGGGGCCCTTCTACCCCAAC 60
        QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyAspLeuLeuProPheLeu 40
        DB 61 CGCCTCTTCGATCAGTCTCGAGAGAAATGTTTGACTAGCCTCTTTCCTTTCTCG 120
        QY 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGly 60
        DB 121 TCCTCCACCGTCAGTCCCTACTACAGGCACAGCCCTCTCCGAGGATTCATGGACTCCGGC 180
        QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
        DB 181 ATTTCTGAGTGCGTTCAGATCGCATCGTTTACCATTAACTTGATGTGAAGCATTTTC 240
        QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyLysHis 100
        DB 241 TCACCCGATGATCTGACTGTGAAGATCTTGATGACTTTTGGAATCCATGGAAAGCAC 300
        QY 101 AsnGluArgGlnAspHisGlyTyIleSerArgGluPheHisArgArgTyArgLeu 120
        DB 301 AGCGAGAGCGAGGATGACCATGGCTACATTTCCCGGAGTTCACCGCGCGCTACCGCTC 360
        QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
        DB 361 CCTTCCAACTGGACCAATCTCTATCAGTGTCTCCTCTGCTGATGGCATTTTGACC 420
        QY 141 PheCysGlyProLysIleGlnThrChlYLeuAspAlaThrHisAlaGluArgAlaIlePro 160
        DB 421 TTCACGCGACCAAAATATGTCAGGCGCTGGACTCCAGCCACAGCGAGAGCCTATCCCT 480
        QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
        DB 481 GTGTCCAGAGAGAAAGCCCACTCTGCCCTCTCTCC 519
        RESULT 13
        LOCUS D86299
        DEFINITION Cynops pyrrhogaster mRNA for new alpha A-crystallin, complete cds.
        ACCESSION D86299
        VERSION D86299.1 GI:19168451
        KEYWORDS Cynops pyrrhogaster (Japanese firebelly newt)
        SOURCE Cynops pyrrhogaster
        ORGANISM Cynops pyrrhogaster
```



```

Db      67 ATGATATACCACTTACAGCACCCCTGGTTCAAGCGCTCCCTGGGGCCCTTCTACCCCAAC 126
Qy      21 ArgLeuPheAspGlnPhePheGlyLeuPheGlyLeuPheGlyLeuPheGlyLeuPheGlyLeu 40
Db      127 CGCCTGTTGACAGGCTTTGGTGAGGAATGTTGACTTGTGACCTGTCCCTTCATG 186
Qy      41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db      187 TCCTCCACATCAGCCCTACTACAGCAGAACCTCTCCAGAGGATACCTGGACTCTGGG 246
Qy      61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPhe 80
Db      247 ATCTCTGAGGTGGTTCAGACCGTTCAGACCGTTCGACCGTTTGTGATTAACCTGGATGTGAAGCATTC 306
Qy      81 SerProGluAspLeuThrValIysValGlnAspAspPheValGluIleHisGlyIleHis 100
Db      307 TCCCTGAGATCTGAGCGTCAAGTCCATGATGACTTTGTAGAGATTTCATGGGAACAC 366
Qy      101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
Db      367 AACGAGCAGCAGGATGATCAGGATACATCTCCGAGGAATTCATCGCGGATATCGTCTT 426
Qy      121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db      427 CTTTCCAATATGATGATCAGAACTCTGTGAGCTGCACTCTGTCTGCGGACGGATCTCACT 486
Qy      141 PheCysGlyProLysIleGlnThrClyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db      487 TTATTCGGTCCCAACTCCAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 546
Qy      161 ValSerArgGluGluLysProThrSerAla 170
Db      547 GTGTCAGGAGGAGGAATCAGGCTCATCC 576

RESULT 16
AJ617724
LOCUS
DEFINITION
  Ornithorhynchus anatinus partial mRNA for alphaA-crystallin (cryaA
  gene).
ACCESSION
  AJ617724.1 GI:50344344
KEYWORDS
  alphaA-crystallin; cryaA gene.
SOURCE
  Ornithorhynchus anatinus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.
REFERENCE
  1
  van Rheede, T. and de Jong, W.W.
  The alpha-crystallins of the platypus Ornithorhynchus anatinus:
  Origin of the alternatively spliced exon alphaAins and implications
  for mammalian phylogeny
  Unpublished
JOURNAL
REFERENCE
  2 (bases 1 to 558)
AUTHORS
  Franck, E.
  Direct Submission
  Submitted (15-DEC-2003) Franck E., Biochemistry, Radboud University
  Nijmegen, 9101, 6500 HB Nijmegen, NETHERLANDS
JOURNAL
  Location/Qualifiers
FEATURES
  source
    1..558
    /organism="Ornithorhynchus anatinus"
    /mol_type="mRNA"
    /db_xref="taxon:9258"
    1..558
    /gene="cryaA"
    /gene="cryaA"
    /codon_start=1
    /product="alphaA-crystallin"
    /protein_id="CAF02100.1"
    /db_xref="GI:50344345"
    /translation="RALGFLYPSRLFDQFGEGLEFYEDLLPFLSTISPPYRQSLFRT
    VLEGISLMAHVFVWHKSHAGNPKNPAKVRSDRDKFVFLFDVGFSPDLTVKVL
    DEFVEIHGKSERQDDHGYISREFHRRYRLPSNVDQASVCSLSLSDGMLTFSGSKVQS

```

```

exon
  NLDSSHSRDSRIPISREEVPTSTPSS"
  157..225
  /gene="cryaA"
  /note="ins exon"
ORIGIN
Alignment Scores:
Pred. No.: 1,39e-72 Length: 558
Score: 728.50 Matches: 139
Percent Similarity: 83.24% Conservative: 15
Best Local Similarity: 75.14% Mismatches: 8
Query Match: 79.53% Indels: 23
DB: 4 Gaps: 1
US-10-657-740-1 (1-173) x AJ617724 (1-558)
Qy 12 ArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhePheGlyGluGlyLeu 31
Db 1 CGTCGGTTGGGGCCTTTATACCCAGCGCCCTGTTTGACCAGTGTTCACCAAGTGTTCGAGAGGGTCTC 60
Qy 32 PheGluTyrAspLeuLeuProPheLeuSerSerThrIleSerProTyrTyrArgGlnSer 51
Db 61 TTCGAGTACGACCTCCCTGCGCTTTCTCTCTCTACCATTCAGTCCTTATTACCGGAGTCT 120
Qy 52 LeuPheArgThrValLeuAspSerGlyIleSerGlu----- 63
Db 121 CTGTTCCGACCGCTCTGGAGTCGGGCAATTCGAGCTCATGGCCCATGTATGTTGTA 180
Qy 64 -----ValArgSerAspArg 68
Db 181 ATGCACAAATCACATGCTGGAAATCCCAAGAAACAACCTGCGCAAGGTCCGATCTGACCGG 240
Qy 69 AspLysPheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValLys 88
Db 241 GACAAGTTGTCTCTCTCGATGTAAGCATTTCTCTCCGAGATCTCAGCGTGAAG 300
Qy 89 ValGlnAspAspPheValGluIleHisGlyIleHisGlyIleHisGlyIleHisGlyIleHisGly 108
Db 301 GTGTTGATGATGTTGTTGGAGATTCATGGCAAGCACAGTGGAGACAGGATGACCAAGT 360
Qy 109 TyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnValAspGlnSerAla 128
Db 361 TACATTTCCCGGAATTCATCGCGGTACCGCTGCTTCCACACGTGGAGACAGGCTCA 420
Qy 129 LeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThr 148
Db 421 GTCTCTGCTCCCTGCTCTCTGATGGCATGTTGACCTTCTCTGCTCCCAAGTCCAAATCC 480
Qy 149 GlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThr 168
Db 481 AACCTGGATTCACGCCACAGTACCGATCCATCCCTATCTCCCGGAGGAAGTGCACACC 540
Qy 169 SerAlaProSerSer 173
Db 541 TCAACTCCCTCTCTCT 555
RESULT 17
AJ617726
LOCUS
DEFINITION
  Sphenodon punctatus partial mRNA for alphaA-crystallin (cryaA
  gene).
ACCESSION
  AJ617726.1 GI:50344348
KEYWORDS
  alphaA-crystallin; cryaA gene.
SOURCE
  Sphenodon punctatus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Lepidosauria; Sphenodontia; Sphenodontidae; Sphenodon.
REFERENCE
  1
  van Rheede, T. and de Jong, W.W.
  The alpha-crystallins of the platypus Ornithorhynchus anatinus:
  Origin of the alternatively spliced exon alphaAins and implications
  for mammalian phylogeny

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 477)
AUTHORS Franck, E.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2003) Franck E., Biochemistry, Radbound University
Nijmegen, 9101, 6500 HB Nijmegen, NETHERLANDS
FEATURES Location/Qualifiers
source 1..477
/organism="Sphenodon punctatus"
/mol_type="mRNA"
/db_xref="taxon:8508"
gene 1..477
/genes="cryaa"
<1..477
/genes="cryaa"
CDS
/codon_start=1
/product="alphaA-crystallin"
/protein_id="CAP02102.1"
/db_xref="GI:50344349"
/translation="ELIPSLRFDQFFGEGLEFYDILLPLFSSTISPIYRQSPFRFTVLES
GISEVRSRDKFTIQLDVHFKFSPDLKSVKIIDDFVEIHGKNERQVDHGYISREPHRR
YRLPSNVDSQSAITCSLSDGDMWTFGAPKQSNMPSHSRPIPVSRREKPTGAPSS"
ORIGIN
Alignment Scores:
Pred. No.: 3,79e-71 Length: 477
Score: 715.00 Matches: 133
Percent Similarity: 91.14% Conservative: 11
Best Local Similarity: 84.18% Mismatches: 14
Query Match: 78.06% Indels: 0
DB: 5 Gaps: 0
US-10-657-740-1 (1-173) x AJ617726 (1-477)
QY 16 ProPheTyrProSerArgLeuPheAspGlnPheGlyGluGlyPheGluTyrAsp 35
DB 1 CCCCTAATTCAGACCGGTTGTTGACACCAAGTTTGGAGAGGCTTTTTCAGTATGAT 60
QY 36 LeuLeuProPheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThr 55
DB 61 CTCCTGGCTTGTCTCTCCACCATCAGCCCTACTACAGCAATCTTCTTCGACCC 120
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspPheValIlePheLeu 75
DB 121 GTCTGGAATCAGGCATTTTCAGAGGTGAGATCTGACCGGACCAAGTTTACATCTCCNG 180
QY 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGlu 95
DB 181 GATGTAACACCTCTCTCTGAAGATTTGAGTGTGAAGATTTATGATGACTTTGTGGAA 240
QY 96 IleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 241 ATCCACGGCAGACACATGAGACAGAGTAGACCTGATGCTATCTCCCGCAATTCAC 300
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 301 CGCAGGTACCGCTGCTCCATGTTGACCAATCGCCATCATCTGCTGCTGCTGCT 360
QY 136 AspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 361 GACGGCATGATGACCTTCTCTGCCCCCAAGGTCCAGCTTAAATGACCCGCCACACAGC 420
QY 156 GluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
DB 421 GAGAGACCATTCCTGTATCCCGTGAAGAGAGCCACCTCGCGCGCTCTCTCC 474
RESULT 18
AY035778
LOCUS AY035778 730 bp mRNA linear VRT 21-JUN-2002
DEFINITION Danio rerio alpha A crystallin mRNA, complete cds.
ACCESSION AY035778
VERSION AY035778.1 GI:18266458
KEYWORDS

SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 730)
AUTHORS Runkle,S., Hill,J., Kantorow,M., Horwitz,J. and Posner,M.
TITLE Sequence and spatial expression of zebrafish (Danio rerio) alphaA-crystallin
JOURNAL Mol. Vis. 8, 45-50 (2002)
MEDLINE 21923047
PUBMED 11925526
REFERENCE 2 (bases 1 to 730)
AUTHORS Runkle,S., Hill,J., Kantorow,M., Horwitz,J. and Posner,M.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Biology, Ashland University, 401 College
Avenue, Ashland, OH 44805, USA
FEATURES Location/Qualifiers
source 1..730
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/tissue_type="lens"
95..615
/note="Small heat shock protein"
/codon_start=1
/product="alpha A crystallin"
/protein_id="AAK61363.1"
/db_xref="GI:18266459"
/translation="MDIALQHPWFRTLGYPTRLFDQFFGEGLEFYDILLPLFPPTSTVSP
YRHSLEFRNLDSSNGVSRDKFTVLDVHFKFSPDLKSVKIIDDFVEIHGKNERQVDHGYISREPHRR
YRLPSNVDSQSAITCSLSDGDMWTFGAPKQSNMPSHSRPIPVSRREKPTGADGRDRTIP
VTREDKSNSSGSSS"
CDS
Alignment Scores:
Pred. No.: 1,51e-68 Length: 730
Score: 694.00 Matches: 127
Percent Similarity: 85.80% Conservative: 24
Best Local Similarity: 72.16% Mismatches: 19
Query Match: 75.76% Indels: 6
DB: 5 Gaps: 3
US-10-657-740-1 (1-173) x AY035778 (1-730)
QY 1 MetAspValThrIleGlnHisProThrPheLysArgThrLeuGlyProPheTyrProSer 20
DB 95 ATGGATATTGGATCCACACCCCTTGGTTTCAGACGCACACTGGGC-----TACCCACC 148
QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
DB 149 CGACTCTTTGATCAGTCTTTGGAGAGGCTGTTGATATGACTATTCCCTTCACC 208
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp----- 58
DB 209 ACTTCAACTGTCAGCCCTTACTATCGACACTCCTCTTCGCAACACTCTCGACTCTCTCC 268
QY 59 ---SerGlyIleSerGluValArgSerAspArgAspPheValIlePheLeuAspVal 77
DB 269 AACTCGGTGTCTCTGAGGTGAGGCTCAGACAGAGAAAATTTACAGTTTATCTGATGTG 328
QY 78 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 97
DB 329 AAACATTTCTCTCTGATGAGCTCAGTGTCAAGGTGACAGATGATATGTGGAGATCCAG 388
QY 98 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 117
DB 389 GGCAAGCATGGAGAAAGACAGATGATCATGGCTACATCTCCGTCGAGTTCATCGTCGC 448
QY 118 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
DB 449 TACCGCTGCTCTTCCAAATGTGACAGCAGTCTGCACACTGACACTGTCCTGCTGATGCG 508

QY 138 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
 Db 509 CGTCTCACTCTTTGGACCCAG--ACAAGTGCATAGATGCTGGCGGTGAGATGCC 565
 QY 158 AlaIleProValSerArgGluGlyProThrSerAlaProSerSer 173
 Db 566 ACCATCCCTGTTTACCGCGGAGCAAGACAACTCAGGCTCTCTCTCC 613

RESULT 19

BC083177 1430 bp mRNA linear VRT 28-SEP-2004
 LOCUS Danio rerio cDNA clone MGC:92036 IMAGE:7045051, complete cds.
 DEFINITION BC083177
 VERSION BC083177.1 GI:52789216
 KEYWORDS MGC.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS 1 (bases 1 to 1430)
 Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.E.,
 Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 12477932
 2 (bases 1 to 1430)
 Director MGC Project.
 Direct Submission
 Submitted (24-SEP-2004) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK

COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs@mail.nih.gov
 Tissue Procurement: Len Zon, Harvard
 cDNA Library Preparation: Open Biosystems
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Akhter,N., Avelle,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granice,S., Guan,X., Gupta,J., Haghighi,P.,
 Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Loric,P., Legaspi,R.,
 Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
 Tsourgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 178 Row: p Column: 11

This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 23308654.

FEATURES

source
 1..1430
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="MGC:92036 IMAGE:7045051"
 /tissue_type="Whole, adult zebrafish"
 /clone_lib="NIH_ZGC_10"
 /lab_host="DH10B"
 /note="Vector: pExpress1"
 106..627
 /codon_start=1
 /product="Unknown (protein for MGC:92036)"
 /protein_id="AAH83177.1"
 /db_xref="GI:52789217"
 /translation="MDIAIQHFWFRRTLGYPTRLDPQFFBGLFYDLPPTTSTVSP
 YRHSLEFRNILDSSNSGVSEVRSEKFTVLDVKHFSDELVSRTDYYEIQKHG
 BRQDDHGYISREFRHRRLPSNVDSAITCTLSADGLLTLCGPKTSGIDAGRGDRTP
 VTREDKSNSSGSS"

CDS

Alignment Scores:
 Pred. No.: 3,55e-68 Length: 1430
 Score: 694.00 Matches: 127
 Percent Similarity: 85.80% Conservative: 24
 Best Local Similarity: 72.16% Mismatches: 19
 Query Match: 75.76% Indels: 6
 DB: 5 Gaps: 3

US-10-657-740-1 (1-173) x BC083177 (1-1430)

QY 1 MetAspValThrIleGlnHisProThrPheLysArgThrLeuGlyProPheTyrProSer 20
 Db 106 ATGGATATTGGGATCCAAACACCTTGGTTCAGACGCACACTGGGC-----TACCCACC 159
 QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 160 CGACTCTTTGATCAGTTCTTTGGAGAGGCGCTGTCGATTATGACCTATTCCCTTCACC 219
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp----- 58
 Db 220 ACTTCAACTGTGAGCCCTTACTATCGACACTCACTCTTCGCAACACTCTGGACTCTCTCC 279
 QY 59 ---SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 77
 Db 280 AACTCAGGTGTCCTGAGTGAGGTGCTGACAGAGAAATTTACAGTTTACCTTGGATGTG 339
 QY 78 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 97
 Db 340 AAACATTTCTCTCTGATGAGCTCAGTGTCAAGGTGACAGATGACTATGTGGAGATCCAG 399
 QY 98 GlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArg 117
 Db 400 GGCAAGCATGGAGAAAGACAGATGATCATGGCTACATCTCCCGTGGTTCATCCATGCCGC 459
 QY 118 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
 Db 460 TACCOCCTGCTCTTCAAGTGGACAGTCTGTCATCACCCTGACACTGTCTGCTGATGCC 519
 QY 138 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
 Db 520 CTGCTCACTCTTTGTGGACCCCAAG--ACAAGTGCATAGATGCTGGCGGTGAGATCCG 576
 QY 158 AlaIleProValSerArgGluGlyProThrSerAlaProSerSer 173
 Db 577 ACCATCCCTGTTTACCGCGGAGCAAGACAACTCGGGCTCTTCTCTCC 624

RESULT 20

TSU31938 448 bp mRNA linear VRT 25-SEP-1996
 LOCUS TSU31938
 DEFINITION Trachemys scripta elegans alpha A-crystallin mRNA, partial cds.

```

ACCESSION      U31938
VERSION         U31938.1
KEYWORDS        GI:1223846
SOURCE          Trachemys scripta elegans
ORGANISM        Trachemys scripta elegans
REFERENCE       1 Trachemys scripta elegans
AUTHORS         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Testudines; Cryptodira; Testudinoidae; Emydidae; Trachemys.
TITLE           1 (bases 301 to 446)
                Hedges, S.B., Simmons, M.D., van Dijk, M.A., Caspers, G.J., de
                Jong, W.W. and Sibley, C.G.
                Phylogenetic relationships of the hoatzin, an enigmatic South
                American bird
JOURNAL         Proc. Natl. Acad. Sci. U.S.A. 92 (25), 11662-11665 (1995)
MEDLINE         96102174
PUBMED         8524824
REFERENCE       2 (bases 1 to 448)
AUTHORS         Caspers, G.J.; Reinders, G.J., Leunissen, J.A., Wattel, J. and de
                Jong, W.W.
TITLE           Protein sequences indicate that turtles branched off from the
                amniote tree after mammals
JOURNAL         J. Mol. Evol. 42 (5), 580-586 (1996)
MEDLINE         96259313
PUBMED         8662010
REFERENCE       3 (bases 1 to 448)
AUTHORS         de Jong, W.W.
TITLE           Direct Submission
                Submitted (20-JUL-1995) Wilfried W. de Jong, Biochemistry,
                University of Nijmegen, P.O.Box 9101, 6500 HB Nijmegen, The
                Netherlands
FEATURES        Location/Qualifiers
                 1..448
                 /organism="Trachemys scripta elegans"
                 /mol_type="mRNA"
                 /sub_species="elegans"
                 /db_xref="taxon:31138"
                 /tissue_type="lens"
                 <1..448
                 /note="related to small heat shock proteins"
                 /codon_start=2
                 /product="alpha A-crystallin"
                 /protein_id="AA08829.1"
                 /db_xref="GI:1223847"
                 /translation="RALGFLPSPRLFDQVLGSLFDYLLPFFSSTISPYVNHSLFRT
                 VLESGVSEVRDRKFTLLDVKHPSFEDLSVKVIMDFVEIHGKNERQDDHGYSIRE
                 PHRRYRLPSNVDSQAITSLSADGMLTTPAPKVQSNMDTSYSERPIP"
CDS             6..7e-67 Length: 448
                Score: 677.00 Matches: 123
                Percent Similarity: 91.95% Conservative: 14
                Best Local Similarity: 82.55% Mismatches: 12
                Query Match: 73.91% Indels: 0
                DB: 5 Gaps: 0
ORIGIN
US-10-657-740-1 (1-173) x TSU31938 (1-448)
QY 12 ArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPheGlyGlyLeu 31
DB 2 CGAGTCTTGGACCTTATTCCAGCCGCTTTGTTGACCATGATCTCGGAGAGGCTCTT 61
QY 32 PheGluTyrAspLeuLeuPheProPheLeuSerSerThrIleSerProTyrTyrArgGlnSer 51
DB 62 TTCGATTGATCTCTGCTTTTCTTCACCATCAGCCCTATTACAGGCACCT 121
QY 52 LeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspLysPhe 71
DB 122 CTCCTCCGACCGCTTCTGGAATCAGGCATTTTCAGAGGTGAGTCTGACCGGACAGTTT 181
QY 72 ValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAsp 91
DB 182 ACAATCTCTCGATGTAACACACTCTCTCCCGAAGATCTCGATGTGAAGATTATGGAT 241
QY 92 AspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSer 111
DB 242 GACTTTGTGGAATCCATCGCAAGCACAATGAGAGACAGGACCATGCTACATTTCC 301
QY 112 ArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCys 131
DB 302 CGCGAATTCACGGCAGATACCGCTGCTTCACAGTGGACCAATCTGCCATCAGCTGC 361
QY 132 SerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAsp 151
DB 362 TCCTCTGCTGCTGATGGCATGCTGACTTTCTCTGCGCCAAAGTCCAGTCCACATGGAC 421
QY 152 AlaThrHisAlaGluArgAlaIlePro 160
DB 422 ACCAGCTATAGCGAGAGACCCCATTCCT 448
RESULT 21
AJ617727
LOCUS          Lygodactylus picturatus partial mRNA for alphaA-crystallin (cryaA
DEFINITION     gene).
ACCESSION      AJ617727
VERSION        AJ617727.1
KEYWORDS       alphaA-crystallin; cryaA gene.
SOURCE         Lygodactylus picturatus
ORGANISM       Lygodactylus picturatus
REFERENCE       1
AUTHORS         van Rheede, T. and de Jong, W.W.
TITLE           The alphaA-crystallins of the platypus Ornithorhynchus anatinus:
                Origin of the alternatively spliced exon alphaAins and implications
                for mammalian phylogeny
JOURNAL         Unpublished
REFERENCE       2 (bases 1 to 447)
AUTHORS         Franck, E.
TITLE           Direct Submission
                Submitted (16-DEC-2003) Franck E., Biochemistry, Radboud University
                Nijmegen, 9101, 6500 HB Nijmegen, NETHERLANDS
FEATURES        Location/Qualifiers
                 1..447
                 /organism="Lygodactylus picturatus"
                 /mol_type="mRNA"
                 /db_xref="taxon:57064"
                 1..447
                 /gene="cryaA"
                 <1..447
                 /gene="cryaA"
                 /codon_start=1
                 /product="alphaA-crystallin"
                 /protein_id="CAF02103.1"
                 /db_xref="GI:50344351"
                 /translation="RAFGFLIPSRFDQFFGGLLDYLLPLFSSTISPYVQSILFRT
                 VLESGVSEVRDRKFTLLDVKHPSFEDLSVKVIMDFVEIHGKNERQDDHGYSIRE
                 PHRRYRLPSNVDSQAITSLSADGMLTTPAPKVQSNMDLSHERPIP"
ORIGIN
Alignment Scores:
Pred. No.: 1.46e-66 Length: 447
Score: 674.00 Matches: 125
Percent Similarity: 89.93% Conservative: 9
Best Local Similarity: 83.89% Mismatches: 15
Query Match: 73.58% Indels: 0
DB: 5 Gaps: 0
US-10-657-740-1 (1-173) x AJ617727 (1-447)
QY 12 ArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPheGlyGlyLeu 31
DB 1 CGGCGTTTGGCCCTTAATTCAGACCGCTTTGTTTCGACCATGCTTTTGGAGAGGCTCTT 60
QY 32 PheGluTyrAspLeuLeuPheProPheLeuSerSerThrIleSerProTyrTyrArgGlnSer 51

```

[illegible]

```

Alignment Scores:
Pred. No.: 3,298-65 Length: 448
Score: 662.00 Matches: 122
Percent Similarity: 89.93% Conservative: 12
Best Local Similarity: 81.88% Mismatches: 15
Query Match: 72.27% Indels: 0
DB: 5 Gaps: 0

US-10-657-740-1 (1-173) x APRAAC (1-448)

QY 12 ArgThrLeuGlyProPheTyProSerArgLeuPheAspGlnPhePheGlyGluGlyLeu 31
Db 2 CGCGCTCGGAGACCCCTGATCCAGCCGTTGTCGACCAAGTTTTCGGAGAGGCTCTC 61
QY 32 PheGluTyArgLeuLeuProPheLeuSerSerThrIleSerProTyTyArgGlnSer 51
Db 62 CTGGAGTATGACCTCGCTGTTGTTCTCTCCATCATCAGCCCTACTACCGGCAGTCC 121
QY 52 LeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspLysPhe 71
Db 122 CTCCTCCGACGCTGCTGGAGTCGGCATTTTCAGAGGTGAGTCTTGACCGGACAAAGTTT 181
QY 72 ValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIysValGlnAsp 91
Db 182 ACGATCATCTGGATGTAACACTTCTCTCTGAAGATCTGAGCTGAAGATTATCGAT 241
QY 92 AspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyIleSer 111
Db 242 GACTTTGTGGAATCCATGCAAGCACACATGAAAGACAGGATGACCAAGGCTACATCTCC 301
QY 112 ArgGluPheHisArgArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCys 131
Db 302 CGTGAGTTTACCGCGGTACCGCTGCCCGCCCAACGTCGACCACTGACCTGACCTGTC 361
QY 132 SerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAsp 151
Db 362 TCCCTCTCCGCGACGCGATGCTGACCTTCTCAGGCCCCCAAGGTCCTCCACATGAC 421
QY 152 AlaThrHisAlaGluArgAlaIlePro 160
Db 422 CCCACCCACGAGAGGCCCATCCCC 448

RESULT 24
CLRAAC 448 bp mRNA linear VRT 09-SEP-2004
LOCUS C.livia mRNA for alpha-A-crystallin.
DEFINITION X96593
ACCESSION X96593
VERSION X96593.1 GI:1945730
KEYWORDS alpha-A-crystallin.
SOURCE Columbia livia (domestic pigeon)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
1
Caspers,G.J., Uit de Weerd,D., Wattel,J. and de Jong,W.W.
AUTHORS alpha-Crystallin sequences support a galliform/anseriform clade
JOURNAL Mol. Phylogenet. Evol. 7 (2), 185-188 (1997)
MEDLINE 97271643
PUBMED 9126559
REFERENCE 2 (bases 1 to 448)
Caspers,G.J.
AUTHORS Direct Submission
TITLE Submitted (13-MAR-1996) G.J. Caspers, University of Nijmegen,
JOURNAL Department of Biochemistry, P.O.Box 9101, NL- 6500 HB Nijmegen,
NETHERLANDS
FEATURES Location/Qualifiers
source 1..448
/organism="Columba livia"
/mol_type="mRNA"
/db_xref="taxon:9932"
/tissue_type="lens"
<!.>448
/codon_start=2
CDS

```

```

/product="alpha-A-crystallin"
/protein_id="CAA65411.1"
/db_xref="GI:1945731"
/db_xref="GOA:O12988"
/db_xref="UniProt/Swiss-Prot:O12988"
/translation="RALGPLIPSLRFDQFEGLELLELLPWFSSSTISPYRQSLFRS
VLESGSEVRSREKFTIMLDVKHFSPELDSVKTIIDDFEIHGKRSEKQDDHGYISRE
FHRRYRUPANVDQAAITCSLSNDGMLTSPGPKVPANMDASHGERP"

ORIGIN
Alignment Scores:
Pred. No.: 3,398-64 Length: 448
Score: 653.00 Matches: 120
Percent Similarity: 89.93% Conservative: 14
Best Local Similarity: 80.54% Mismatches: 15
Query Match: 71.29% Indels: 0
DB: 5 Gaps: 0

US-10-657-740-1 (1-173) x CLRAAC (1-448)

QY 12 ArgThrLeuGlyProPheTyProSerArgLeuPheAspGlnPhePheGlyGluGlyLeu 31
Db 2 CGTGCTCGGACCCCTCATTCACAGCCGTTGTTTCGACCAAGTTTTCGGAGAGGCTCTC 61
QY 32 PheGluTyArgLeuLeuProPheLeuSerSerThrIleSerProTyTyArgGlnSer 51
Db 62 CTCGAGTACGATCTCTGCTGTTGTTCTCTCCATCATCAGCCCTACTACAGGCAGTCC 121
QY 52 LeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspLysPhe 71
Db 122 CTCCTCCGACGCTGCTGGAGTCGGCATTTTCAGAGGTGAGTCTTGACCGGAAAGTTT 181
QY 72 ValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIysValGlnAsp 91
Db 182 ACAATCATCTGGATGTAACACTTCTCTCCGAGACCTGAGTGTGAAGATTATTGAC 241
QY 92 AspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyIleSer 111
Db 242 GACTTTGTGGAATCCATGCAAGCACACATGAAAGACAGGATGACCAAGGCTACATCTCC 301
QY 112 ArgGluPheHisArgArgTyArgLeuProSerAsnValAspGlnSerAlaLeuSerCys 131
Db 302 CGCGAATTCACCGCGGTACCGCTGCCCGCCCAACGTCGACCACTGACCTGACCTGTC 361
QY 132 SerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAsp 151
Db 362 TCCCTGTCCACGACGCGATGCTGACCTTCTCGGCCCCCAAGGTCCTCCACATGAC 421
QY 152 AlaThrHisAlaGluArgAlaIlePro 160
Db 422 GCCAGCCACGCGAGAGGCCCATCCCC 448

RESULT 25
AY007972 676 bp mRNA linear VRT 23-OCT-2000
LOCUS Clarias fuscus alpha-A crystallin mRNA, complete cds.
DEFINITION Clarias fuscus alpha-A crystallin mRNA, complete cds.
ACCESSION AY007972
VERSION AY007972.1 GI:10946518
KEYWORDS
SOURCE Clarias fuscus (whitespotted clarias)
ORGANISM Clarias fuscus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Clariidae; Clarias.
1 (bases 1 to 676)
Chiou,S.-H. and Yu,C.-M.
AUTHORS Alpha crystallin of catfish eye lenses: cDNA and genomic analysis of
TITLE alpha-A and alpha-B
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 676)
Chiou,S.-H. and Yu,C.-M.
AUTHORS Direct Submission
TITLE Submitted (15-SEP-2000) Institute of Biological Chemistry, Academia
JOURNAL

```

FEATURES
source Sinica, P.O. Box 23-106, Taipei, Taiwan 10098, Republic of China
Location/Qualifiers
1..676
/organism="Clarias fuscus"
/mol_type="mRNA"
/db_xref="taxon:33541"
CDS
136..657
/codon_start=1
/product="alpha-A crystallin"
/protein_id="AAG23866.1"
/db_xref="GI:10946519"
SYRSLRFLDSSNSGISEVRSRDRPMVYLDVKHSPELRVKVVADYVEIHKHG
ERQDDHGYSREPHRRYLPNSVDQAATCTLSADGLLSFCGPKTGSGKYGRDRTIP
VTRDDKPNFAASS"

ORIGIN

Alignment Scores:
Pred. No.: 5,728-64 Length: 676
Score: 653.00 Matches: 124
Percent Similarity: 83.52% Conservative: 23
Best Local Similarity: 70.45% Mismatches: 23
Query Match: 71.29% Indels: 6
DB: 5 Gaps: 3
US-10-657-740-1 (1-173) x AY007972 (1-676)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyProSer 20
Db 136 ATGGATATTGCCATCAATCCCGTGGTTCAGACGTACCCCTGAGC-----AATCCCTCT 189
QY 21 ArgLeuAspGlnPhePheGlyGluGlyLeuPheGluTyArgAspLeuLeuProPheLeu 40
Db 190 CGACTCTTTGACCAATCTTTGGAGAAGGTCTGCTTGATCATGACTCTCCCTTTACT 249
QY 41 SerSerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAsp----- 58
Db 250 GCCTCCACCATCAGCCCTCTCCACGACACTCTCTTCCGAGTGTTCGACTCCCTCC 309
QY 59 ---SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 77
Db 310 AACTCTGGCATTTCCGAGGTGAGGTCTGCACAGACAGATTCATGGTTTACCTGGATGTG 369
QY 78 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 97
Db 370 AAGCACTTCTCCCGAGAGAACTCAGAGTTAAGGTAGCAGACTATGTGGAGATTCTAT 429
QY 98 GlyLysHisAsnGluArgGlnAspAspHisGlyTyIleSerArgGluPheHisArgArg 117
Db 430 GGCAGCATGGAGAAAGACAGACGACCATGGCTACATCTCAGCGGAGTTCACCGCGC 489
QY 118 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
Db 490 TACCCTGTACCTCCCAAGCTCGATCAGGCGGCATCACTTGTACGCTGTACGCTGATGC 549
QY 138 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
Db 550 CTGCTCAGCTTCTCGGGCCCCAANA---ACAGCGGATCAAAATACGGCGGTGGAGATCGC 606
QY 158 AlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 607 ACCATTCTGTCTACCCGGGATGACAAACCAACCCCTGTCTCTCTCT 654

RESULT 26
LOCUS RTRCYA 765 bp mRNA linear VRT 06-JUL-1989
DEFINITION Frog mRNA fragment for alpha-A2-crystallin.
ACCESSION X00716 X00058
VERSION X00716.1 GI:54292
KEYWORDS alpha-crystallin; crystallin.
SOURCE Rana temporaria (common frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS Tomarev,S.I., Zinov'eva,R.D., Dolgilevich,S.M., Krayev,A.S., Skryabin,K.G. and Gause,G.G. Jr.
TITLE The absence of the long 3'-non-translated region in mRNA coding for eye lens alpha A2-crystallin of the frog (Rana temporaria)
JOURNAL FEBS Lett. 162 (1), 47-51 (1983)
MEDLINE 84005173
PUBMED 6604666
REFERENCE
AUTHORS Tomarev,S.I., Zinov'eva,R.D., Kraev,A.S., Skryabin,K.G. and Gause,G.G.
TITLE Primary structure of cloned cDNA coding alpha-A2 crystallin of the eye lens of the frog Rana temporaria
JOURNAL Dokl. Biochem. 271, 277-280 (1984)
COMMENT Data kindly reviewed (05-11-1985) by S.I. Tomarev
Sequence 1 to 185 is complementary to sequence 553 to 369 and probably a cloning artefact; amino terminal amino acids are missing.
FEATURES
source Location/Qualifiers
1..765
/organism="Rana temporaria"
/mol_type="mRNA"
/db_xref="taxon:8407"
misc_feature 1..185
/note="artefactual sequence"
CDS 186..632
/note="unnamed protein product; Protein sequence is in conflict with the conceptual translation; alpha-A2 crystalline (aa 25 to 173)"
/codon_start=1
/protein_id="CAA25308.1"
/db_xref="GI:1334774"
/db_xref="GOA:P02508"
/db_xref="UniProt/Swiss-Prot:P02508"
/translation="QVFGEMFDYDLFPFLTSTVSPHYRGLLGRFMDSGISEVRSRDRPTINLDVKHFSPPDLTKLDDFVEIHGKHSERQDDHGYISREPHRRYLRPSNLQDSSISCSLSADGILLTFSGPKMNSLVSSHSEPIPVSRREKFTSAPSS"
misc_feature 744..749
/note="polyadenylation signal"
polyA_site 765
/note="polyadenylation site"
ORIGIN
Alignment Scores:
Pred. No.: 2,02e-61 Length: 765
Score: 631.00 Matches: 116
Percent Similarity: 88.82% Conservative: 19
Best Local Similarity: 76.32% Mismatches: 17
Query Match: 68.89% Indels: 0
DB: 5 Gaps: 0
US-10-657-740-1 (1-173) x RTRCYA (1-765)
QY 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyArgPheLeuProPheLeuSer 41
Db 177 CTTACAGCTCAGGCTTCGGAGAGGGGATGTTTGACTAGGACCTCTCCCTTCTCTGACC 236
QY 42 SerThrIleSerProTyTyArgGlnSerLeuPheArgThrValLeuAspSerGlyIle 61
Db 237 TCCACCGTCAGCCGCCACTACCGGCACCGCCTCTCTCCGAGGATTCATGAGCTCCGGCATC 296
QY 62 SerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSer 81
Db 297 TCTGAGGTCGCTCAGATCGCGGATCGTTTACCATTAACTTGAGCTGAAGATTCTCTCC 356
QY 82 ProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsn 101
Db 357 CCCGATGATCTGACTGTGAGATTCTGGATGATCTTTGTGGAATCCATGGAAGCACAGC 416
QY 102 GluArgGlnAspAspHisGlyTyIleSerArgGluPheHisArgArgTyThrGluPro 121

```

Db 417 GAGAGCAGGATGACCATGGCTACATATCCCGGAGTTCCACCGCGCTACCGCTCCCC 476
Qy 122 SerAenValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 141
Db 477 TCCAACTGGACCAATCTCTATCAGCTGCTCACTCTGCGGATGCGATTCTGACCTTC 536
Qy 142 CysGlyProLyIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProVal 161
Db 537 AGCGACCAAAATGATGTCAAACCTGGTCTCCAGCCACAGGAGAGGCCCATCTCCGTG 596
Qy 162 SerArgGluGluLysProThrSerAlaProSerSer 173
Db 597 TCCAGAGAGAAAGCCACCTCTGCCCTCTCTCC 632

RESULT 27
RNACRY
LOCUS
DEFINITION Partial sequence of messenger RNA for rat alpha-a2-crystallin.
ACCESSION V01219 J00715
VERSION V01219.1 GI:55598
KEYWORDS complementary DNA; crystallin.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 885)
AUTHORS Moormann,R.J., van der Velden,H.M., Dodemont,H.J., Andreoli,P.M.,
Bloemendal,H. and Schoenmakers,J.G.
TITLE An unusually long non-coding region in rat lens alpha-crystallin.
JOURNAL Nucleic Acids Res. 9 (19), 4813-4822 (1981)
MEDLINE 82081811
PUBMED 6171772
FEATURES
source
Location/Qualifiers
1..885
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
<1..885
<1..368
/codon_start=3
/product="crystallin"
/protein_id="CAA24530.1"
/db_xref="GI:809074"
/translation="FRTVLDGSGISVRSRDKFIFLDVKHFSPEDLTVKVLDFVEI
HGKNERQDDHGYSIREFRRYRLPSNVDSALSCSLSDGMLTFSGPKVQSGLDAGH
SERALPVSREKPSAPSS"
polyA_site
ORIGIN
885
Alignment Scores:
Pred. No.: 7,83e-57 Length: 885
Score: 591.00 Matches: 113
Percent Similarity: 97.52% Conservative: 5
Best Local Similarity: 93.39% Mismatches: 3
Query Match: 64.52% Indels: 0
Gaps: 0
DB: 10

US-10-657-740-1 (1-173) x RNACRY (1-885)

Qy 53 PheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheVal 72
Db 3 TTCCGCACAGTGTGGATCTCCGGCATCTCTGAGGTCCGATCTGACCGGACAGATTGTC 62
Qy 73 IlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAsp 92
Db 63 ATCTTCTGGATGTGAAGCACTTCTCTCTGAGGACCTCACCGTGAAGGACTTGGAGAT 122
Qy 93 PheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArg 112
Db 123 TTCTGGAGATCCATGGCAACACACACAGAGGAGGATGATCCATGGCTACATTTCCGT 182

```

```

Qy 113 GluPheHisArgArgTyrArgLeuProSerAenValAspGlnSerAlaLeuSerCysSer 132
Db 183 GAAATTCACCGTCGCTACCGTCTGCTTCCAAATGTGGACCACTCCGCCCTCTCTGCTCC 242
Qy 133 LeuSerAlaAspGlyMetLeuThrPheCysGlyProLyIleGlnThrGlyLeuAspAla 152
Db 243 TTGTCGCGGATGGCATGCTGACCTTCTCTGCCCCCAAGGTCAGTCTGGCTTGGATGCT 302
Qy 153 ThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSer 172
Db 303 GCACACAGGAGAGGCCCATTTCCCGTGTCCAGGAGGAGAGGCCAGCTCGGCACCTCG 362
Qy 173 Ser 173
Db 363 TCC 365

RESULT 28
LOCUS
DEFINITION Oryzias latipes, alpha-A-crystallin gene.
ACCESSION AJ000940
VERSION AJ000940.1 GI:3115327
KEYWORDS alpha-A-crystallin gene.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1
AUTHORS Loosli,F., Koster,R.W., Carl,M., Krone,A. and Wittbrodt,J.
TITLE Six3, a medaka homologue of the Drosophila homeobox gene sine
oculus is expressed in the anterior embryonic shield and the
developing eye
JOURNAL Mech. Dev.
REFERENCE 2 (bases 1 to 436)
AUTHORS Loosli,F.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1998) Loosli F., Institute for Human Genetics,
c/o MPI of Biophys. Chem., University of Göttingen, Am Fassberg,
Göttingen, 37077, GERMANY
FEATURES
source
Location/Qualifiers
1..436
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Carolina Biological"
/db_xref="taxon:8090"
1..436
/gene="alpha-A-crystallin"
<1..436
/gene="alpha-A-crystallin"
/codon_start=1
/protein_id="CAA04397.1"
/db_xref="GI:3115328"
/db_xref="GOA:O73919"
/db_xref="UniProt/Swiss-Prot:O73919"
/translation="RLFDQREGMGFDHLLPFTSPITSPFVQSLFENFLDSSNGI
SEVRSRDKFTVHWVKHFSFDELSEVKVLDLDFVSIQKHGQRQDDHGYSIREFRRYR
LPSTVDQSAITCSLSADGLTLTSGPNAGGNGRSDRSIPVCR"
ORIGIN
436
Alignment Scores:
Pred. No.: 6,31e-52 Length: 436
Score: 544.00 Matches: 106
Percent Similarity: 82.88% Conservative: 15
Best Local Similarity: 72.60% Mismatches: 21
Query Match: 59.39% Indels: 4
Gaps: 2
DB: 5

US-10-657-740-1 (1-173) x OJ000940 (1-436)

Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluLeuProPheLeu 40

```

```

Db      1  CGCGTGTTCGACGACGATTTTTCGGGAGGGGATGTTTCGATCAGCAGCCTCTCGCCCTTACC 60
Qy      41  SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp----- 58
Db      61  TCCCCCACCACATAGTCCGCTTACAGACAGTCGCTGTTTCGTAACCTCTCGACTCTCTCA 120
Qy      59  ---SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVal 77
Db      121  AACTCCGGCATTTCCGAGGTGAGTCTGACAGGACAGTTCACGGTTCACCTGGGATGTC 180
Qy      78  LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 97
Db      181  AAGCACTTCTCCCGCATGAGCTCAGCGTGAAGGTGATCGATGACTTGTGTGAGATCCAG 240
Qy      98  GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArg 117
Db      241  GCGAAGCATGGAGAAAGACAGATACCATGGCTACATCTCCCGGAGTTCACCGCGGC 300
Qy      118  TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
Db      301  TACCGCTCCCTCCACTGTGGACCAATCGCCATCAGCTGCTCCTCTGTGCGGATGGA 360
Qy      138  MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
Db      361  CTGCTGACCTGTCTCCGGGCCAAACCCAGCGGGGGCGGGAATGGA---CGCAGCGACGCGC 417
Qy      158  AlalieProValSerArg 163
Db      418  AGCATCCCGCTGTCGCCG 435

RESULT 29
AF029793
LOCUS      Bos taurus alpha B-crystallin (CRYAB) mRNA, complete cds.
DEFINITION
ACCESSION  AF029793
VERSION    AF029793.2 GI:5296002
KEYWORDS
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE  1  (bases 1 to 632)
            Kelley,P.B., Abraham,E.C., Zhao,H.R., Shroff,N.P., Cherian,M. and
            Thomas,J.J.
            Direct Submission
            Submitted (10-OCT-1997) Biochemistry and Molecular Biology, Medical
            College of Georgia, 1120 15th Street, Augusta, GA 30912, USA
            Kelley,P.B., Abraham,E.C., Zhao,H.R., Shroff,N.P., Cherian,M. and
            Thomas,J.J.
            Direct Submission
            Submitted (30-JUN-1999) Biochemistry and Molecular Biology, Medical
            College of Georgia, 1120 15th Street, Augusta, GA 30912, USA
            Sequence update by submitter
            On Jun 30, 1999 this sequence version replaced gi:2760900.
FEATURES             location/Qualifiers
     source           1..632
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
     gene             1..632
                     /gene="CRYAB"
     CDS              1..528
                     /gene="CRYAB"
                     /codon_start=1
                     /product="alpha B-crystallin"
                     /protein_id="BAB95323.2"
                     /db_xref="GI:5296003"
                     /translation="MDIAIHPIWIRPFPHSPRLPDQFFGEHLLSDLPFPASTSL
                     SPFYLRPSFLRAPSFLNDGLSEMRLEKDFSVNLDVKHFSPEELKVKVLGVLEIYVHG
                     KHEERQDHGFIISREIFHRKRIYPADVDPLAITSLSSDGLTVNVPKQASGPRTIP
                     ITREKPAVTAAPKK"

```

```

ORIGIN
Alignment Scores:      1.19e-46      Length:      632
Pred. No.:            499.00      Matches:      99
Score:                75.28%      Conservative: 35
Percent Similarity:   55.62%      Mismatches:  32
Best Local Similarity: 54.48%      Indels:      12
Query Match:          4          Gaps:        6
DB:

US-10-657-740-1 (1-173) x AF029793 (1-632)
Qy      1  MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      1  ATGGATATCGCATCCACCCATCCGCGCCCTTCTTCCCTTCCACTCTCCC 60
Qy      20  SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe. 39
Db      61  AGCGCGCTCTTTTGACCACTTTTTCGGGAGACCTGTTGGAGTCTGATCTCTTCCCA--- 117
Qy      40  LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db      118  GCTTCTACTTCCCTGAGCCCTTCTACCTTCGCGCCCTCATTTCTCGGGCACCACG 177
Qy      56  ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db      178  TGGATTGACACTGGCTCTCAGAGATGCTCTGGAGAGGACAGATTCTCTGTCAACCTG 237
Qy      76  AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db      238  GATGTGAAGCACTTCTCCCGAGAGAACTCAAGGTCAAGGTCTGGGAGATGTGATTGAG 297
Qy      96  IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db      298  GTGATGTCGCAAAACATGAAGAGCGCAGATGATGATGTTTATCTCCCGGAGTTCAC 357
Qy      116  ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db      358  AGGAATATCCGATCCCGAGCTGACGTGGACCTCTCGCCATTACTTCCTCTGCTCT 417
Qy      136  AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db      418  GATGGGTCTCTCACTGTGAATGGACCAAGGAACAG-----GCCTCGCGCCCT 465
Qy      156  GluAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db      466  GAGCGACCATTCATCCATCCCGTGAAGAGAGCGCGCTGTCACTGCGAGCCCCC 519

RESULT 30
OCERYAB
LOCUS      O.cuniculus mRNA for alpha-B-crystallin.
DEFINITION
ACCESSION  X95383
VERSION    X95383.1 GI:1177578
KEYWORDS   alpha-B-crystallin; cryaB gene.
SOURCE     Oryctolagus cuniculus (rabbit)
ORGANISM   Oryctolagus cuniculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE  1
            Krausz,E., Augusteyn,R.C., Quinlan,R.A., Reddan,J.R., Russell,P.,
            Sax,C.M. and Graw,J.
            Expression of Crystallins, Pax6, Filensin, CP49, MIP, and MP20 in
            lens-derived cell lines
            Invest. Ophthalmol. Vis. Sci. 37 (10), 2120-2128 (1996)
            96409169
            8814151
            Graw,J.
            Direct Submission
            Submitted (29-JAN-1996) J. Graw, Institute of Mammalian Genetics,
            GSF-Research Center Neuherberg, Ingolstaedter Landstrasse 1,
            Oberschleissheim, D-85764, FRG

```

```

FEATURES             source
  Location/Qualifiers
    1..548
      /organism="Oryctolagus cuniculus"
      /mol_type="mRNA"
      /sub_species="New Zealand white"
      /db_xref="taxon:9986"
      /tissue_type="lens"
    1..548
      /gene="cryab"
    21..548
      /gene="cryab"
      /codon_start=1
      /product="alpha-B-crystallin"
      /protein_id="CAA64669.1"
      /db_xref="GI:1177579"
      /db_xref="GOA:P41316"
      /translation="MDIAIHHPWIRPPFPSPSRLEFDQPFGEHLLSDLEPTSTSL
      SPFYLRPSPFLAPSWIDTGLSEMRLEKDRFSVNDVVKHFSPEELKVKVLGVDVIEVHG
      KHEERQDHGFISREFHRRKYRIPADVDPLTITSSLSGVLTVNGPRKQVGPRTIP
      ITREKPAVTAAPKK"

ORIGIN
Alignment Scores:
  Pred. No.:      4,728-46      Length:      548
  Score:          493.00      Matches:    98
  Percent Similarity: 74.16%      Conservative: 34
  Best Local Similarity: 55.06%      Mismatches: 34
  Query Match:      53.82%      Indels:    12
  DB:                4          Gaps:        6

US-10-657-740-1 (1-173) x OCRYAB (1-548)

Qy      1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      21 ATGGACATCGCTATCACACCCCTGGATCCGCCGCCCTTCTTCCCTTTCACCTGCC 80

Qy      20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db      81 AGCCGCTCTTGGACAGTCTTCGGAGACACCTGTTGGAGTGTGATCTCTTCCA--- 137

Qy      40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db      138 ACTTCTACTTCCCTGAGCCCTTCTATCTTCGCCACCCCTCATCTCTGGCGCACCCACG 197

Qy      56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspGlyPheValIlePheLeu 75
Db      198 TGGATTGACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTCTCAACCTG 257

Qy      76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db      298 GATGTGAGCACTTCTCCCGAGAGAGCTCAAGTCAAGTGTGGGTGATGTGATTGAG 317

Qy      96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db      318 GTGCACGGCAACATGAGAGCGCCAGGATGAAGTGTTCATCTCTCAGGGAGTTCAC 377

Qy      116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db      378 AGGAAATACCGGATCCAGCTGATGGACCTCTCACCATTACTTCACTTCCCTGTCTCT 437

Qy      136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db      438 GATGGGTCTCTACTGTGTGATGACCAAGAGCA-----GCCCTGGCCCA 485

Qy      156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db      486 GAGCGCACCATTTCCCATACCGGTGAAGAGAGCGCTGTGTCTACTGCAGCCCCC 539

RESULT 31
BT006770
LOCUS
DEFINITION Homo sapiens crystallin, alpha B mRNA, complete cds.

ACCESSION BT006770
VERSION BT006770.1
KEYWORDS GI:30582378
SOURCE FLI CDNA.
ORGANISM Homo sapiens (human)
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 528)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor
vector
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 528)
AUTHORS Kainine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
Phelan,M. and Farmer,A.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
Circle, Palo Alto, CA 94303, USA
COMMENT This CDS clone is a part of a collection of human full length
expression clones generated by BD Biosciences Clontech and the
Harvard Institute of Proteomics. Each CDS has been cloned in two
forms: with and without stop-codon (to allow fusion with C-terminal
tag). The CDS has been directionally cloned using BD In-Fusion(TM)
cloning system between the SalI and HindIII sites of the pDNR-DUAL
vector. Additional sequences in the clone: 'ACC' after SalI site
and before 'ATG' to provide Kozak consensus sequence; 'GG' after
last codon and before HindIII site to maintain reading frame.
Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
  source
    1..528
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="GH00548X1.0"
      /clone_lib="BD Creator(TM) CDS Library derived from MGC
      collection"
      /lab_host="DH5alpha TI resistant"
      /note="vector: pDNR-Dual"
    1..528
      /codon_start=1
      /product="crystallin, alpha B"
      /protein_id="AAP35416.1"
      /db_xref="GI:30582379"
      /translation="MDIAIHHPWIRPPFPSPSRLEFDQPFGEHLLSDLEPTSTSL
      SPFYLRPSPFLAPSWIDTGLSEMRLEKDRFSVNDVVKHFSPEELKVKVLGVDVIEVHG
      KHEERQDHGFISREFHRRKYRIPADVDPLTITSSLSGVLTVNGPRKQVGPRTIP
      ITREKPAVTAAPKK"

ORIGIN
Alignment Scores:
  Pred. No.:      1,278-45      Length:      528
  Score:          489.00      Matches:    97
  Percent Similarity: 73.60%      Conservative: 34
  Best Local Similarity: 54.49%      Mismatches: 35
  Query Match:      53.38%      Indels:    12
  DB:                9          Gaps:        6

US-10-657-740-1 (1-173) x BT006770 (1-528)

Qy      1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      1 ATGGACATCGCTATCACACCCCTGGATCCGCCGCCCTTCTTCCCTTTCACCTGCC 60

Qy      20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db      61 AGCCGCTCTTGGACAGTCTTCGGAGACACCTGTTGGAGTGTGATCTTTTCCG--- 117

Qy      40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db      118 ACGTCTACTTCCCTGAGTCCCTTACCTTCGGCCACCCCTCTCTCTCGGGCACCCAGC 177

```



```

source
1. .537
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 1.3e-45 Length: 537
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x AX937702 (1-537)

QY 1 MetAspValThrIleGlnHisProTrrPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 9 ATGGACATCGCCATCCACACCCCTGGATCCGCGGCCCTTCTTTCTTCCACTCCCC 68
QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 69 AGCCGCTCTTTGACCACTTCTCGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 125
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 126 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCGCACCCCTCTCTCGCGGCACCCAGC 185
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 186 TGGTTTGACACTGGATCTTCAGAGATCGCCTGGAGAGACAGGTCTCTGTGCAACCTG 245
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
DB 246 GATGTGAAGCACTTCTCCCGAGGAACCTCAAAGTTAAGGTGTGGAGATGTGATTGAG 305
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 306 GTGCATGAAACATCATGAAGAGCGCCAGGATGAACATGTTTCATCTCCAGGGAGTTCCAC 365
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 366 AGAATAATACCGATCCAGATGATGATGAGACCTCTCAGACCTCTCAGACCTCTCATCT 425
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 426 GATGGGCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCT 473
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 474 GAGCGCACCATTCCCATCCCGGAGAGAGAGCTGTCTGTCACCGCAGCCCC 527

RESULT 34
AX937703
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .537
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

misc_feature
1. .691
/note="S45630"

ORIGIN
Alignment Scores:
Pred. No.: 1.3e-45 Length: 537
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x AX937703 (1-537)

QY 1 MetAspValThrIleGlnHisProTrrPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 9 ATGGACATCGCCATCCACACCCCTGGATCCGCGGCCCTTCTTTCTTCCACTCCCC 68
QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 69 AGCCGCTCTTTGACCACTTCTCGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 125
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 126 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCGCACCCCTCTCTCGCGGCACCCAGC 185
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 186 TGGTTTGACACTGGATCTTCAGAGATCGCCTGGAGAGACAGGTCTCTGTGCAACCTG 245
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
DB 246 GATGTGAAGCACTTCTCCCGAGGAACCTCAAAGTTAAGGTGTGGAGATGTGATTGAG 305
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 306 GTGCATGAAACATCATGAAGAGCGCCAGGATGAACATGTTTCATCTCCAGGGAGTTCCAC 365
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 366 AGAATAATACCGATCCAGATGATGATGAGACCTCTCAGACCTCTCAGACCTCTCATCT 425
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 426 GATGGGCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCT 473
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 474 GAGCGCACCATTCCCATCCCGGAGAGAGAGCTGTCTGTCACCGCAGCCCC 527

RESULT 35
AX937703
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .691
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

misc_feature
1. .691
/note="S45630"

ORIGIN

```



```

Percent Similarity: 73.60%      Conservative: 34
Best Local Similarity: 54.49%    Mismatches: 35
Query Match: 53.38%             Indels: 12
DB:                               Gaps: 6

US-10-657-740-1 (1-173) x AX333032 (1-691)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGGACATCGCCATCCACACCCCTGGATCCGCCGCCCTCTTCTTCTTCTTCCACTCCCCC 85
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCCCTCTTGGACAGTCTTCGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 142
QY 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55
Db 143 ACGTCTACTTCCCTGAGTCCCTTACCTTCGCCACCCCTCTTCTTCTGGGGCACCACCG 202
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 203 TGGTTTGACACTGACACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 262
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 263 GATGTGAAGCACTTCTCCCGAGAGCACTCAAAGTTAAGGTGTGGGAGATGTGATTGAG 322
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGGAACAACATGAAGAGCCGAGATGAACATGTTTCACTTCAGGGAGTTCCAC 382
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 383 AGGAAATACCGGATCCCGAGTGTAGACCTCTCACCATTACTTCTCATCTGTCATCT 442
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGGTCTCCTCAGTGTGAATGGAGCAAGGAAACAG-----GTCTCTGGCCCT 490
QY 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTCCTCCATCACCCTGGAAGAGAGCCCTGCTGTCCACCGACCCCCC 544

RESULT 39
BD079402
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Old, L.J., Scanlan, M.J., Stockert, E., Gure, A., Chen, Y.T., Gout, I.,
Oghare, M., Obata, Y., Pfreundschuh, M., Tureci, O. and Sahin, U.
Cancer-associated nucleic acids and polypeptides
Patent: JP 2001516009-A 68 25-SEP-2001;
LUDWIG INSTITUTE FOR CANCER RESEARCH
OS Homo sapiens (human)
PN JP 2001516009-A/68
PD 25-SEP-2001
PF 15-JUL-1998 JP 2000503425
PR 17-JUL-1997 US 08/896164,10-OCT-1997 US 60/061599 PR
10-OCT-1997 US 60/061765,10-OCT-1997 US 08/948705 PR
11-OCT-1997 GB 9721697.2,22-JUN-1998 US 09/102322 PI LLOYD
J OLD, MATTHEW J SCANLAN, ELISABETH STOCKERT, ALI GURE, YAO PI
CHEN, IVAN GOUT, MICHAEL O'HARE, YUICHI OBATA, MICHAEL PFREUNDSCHEH, PI
OZLEM TURECI,
PI UGUR SAHIN
PC

```

```

G01N33/574, A61K38/00, A61K39/395, A61K39/395, A61K45/00, A61K48/00, PC
A61P35/00,
PC C07K14/82, C07K16/32, C12N15/09//C07K16/46, C12P21/08, A61K37/02,
PC C12N15/00
CC Cancer-associated nucleic acids and polypeptides. FH Key
Location/Qualifiers
FT source 1. .691
FT location/Qualifiers
1. .691
/organism="Homo sapiens (human)".
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="caxon:9606"

ORIGIN

Alignment Scores: 1.79e-45 Length: 691
Pred. No.: 489.00 Matches: 97
Score: 73.60% Conservative: 34
Percent Similarity: 54.49% Mismatches: 35
Best Local Similarity: 53.38% Indels: 12
Query Match: 6 Gaps: 6
DB:

US-10-657-740-1 (1-173) x BD079402 (1-691)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGGACATCGCCATCCACACCCCTGGATCCGCCGCCCTCTTCTTCTTCTTCCACTCCCCC 85
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCCCTCTTGGACAGTCTTCGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 142
QY 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55
Db 143 ACGTCTACTTCCCTGAGTCCCTTACCTTCGCCACCCCTCTTCTTCTGGGGCACCACCG 202
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 203 TGGTTTGACACTGACACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 262
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 263 GATGTGAAGCACTTCTCCCGAGAGCACTCAAAGTTAAGGTGTGGGAGATGTGATTGAG 322
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGGAACAACATGAAGAGCCGAGATGAACATGTTTCACTTCAGGGAGTTCCAC 382
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 383 AGGAAATACCGGATCCCGAGTGTAGACCTCTCACCATTACTTCTCATCTGTCATCT 442
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGGTCTCCTCAGTGTGAATGGAGCAAGGAAACAG-----GTCTCTGGCCCT 490
QY 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTCCTCCATCACCCTGGAAGAGAGCCCTGCTGTCCACCGACCCCCC 544

RESULT 39
S45630
LOCUS
DEFINITION
alpha B-crystallin-Rosenthal fiber component [human, glioma cell
line, mRNA, 691 nt].
ACCESSION
S45630
VERSION
S45630.1 GI:256398
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 691)
 AUTHORS Iwaki,A., Iwaki,T., Goldman,J.E., Ogomori,K., Tateishi,J. and Sakaki,Y.
 TITLE Accumulation of alpha B-crystallin in brains of patients with Alexander's disease is not due to an abnormality of the 5'-flanking and coding sequence of the genomic DNA
 JOURNAL Neurosci. Lett. 140 (1), 89-92 (1992)
 MEDLINE 93025869
 PUBMED 1407707
 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gisbseq 114959] from the original journal article.
 FEATURES
 Location/Qualifiers
 1..691
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 gene
 1..691
 /gene="alpha B-crystallin"
 CDS
 26..553
 /gene="alpha B-crystallin"
 /note="Rosenthal fiber component"
 /codon_start=1
 /product="alpha B-crystallin"
 /protein_id="AAB23453.1"
 /db_xref="GI:256339"
 /translation="MDIAIHHPWIRRRPFPFPHSRRLFDQFFGHEHLLSDLPPTSTLS
 SPFYLRPSFTLRAPSWEDTGLSEWLEKDRFVNLDVGHFSPSELKVKVLGDVIEVHG
 KHERQDGHGFIISREHFKRYPADVDPLTITSSLSDDGLTVNGPRKQNSGPRTIP
 ITREKPAVTAAPK"

ORIGIN

Alignment Scores:
 Pred. No.: 1,79e-45 Length: 691
 Score: 489.00 Matches: 97
 Percent Similarity: 73.60% Conservative: 34
 Best Local Similarity: 54.49% Mismatches: 35
 Query Match: 53.38% Indels: 12
 DB: 9 Gaps: 6

US-10-657-740-1 (1-173) x S45630 (1-691)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
 Db 26 ATGGACATCGCATCCACCACCCCTGGATCGCGCCCTTCTTCTTCCATCTCCGCC 85
 QY 20 SerArgLeuPheAspGlnPhePheGlyGluPheGluTyrAspLeuLeuProPhe 39
 Db 86 AGCCGCCCTTTGACACAGTTCTTCGGAGACACCTGTTGGAGTCTGTATCTTTCCCG--- 142
 QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg---Thr 55
 Db 143 AGCTTACTTCCCTGAGTCCCTTCTACCTTGGCCACCCCTCTTCTCGGGCACCACG 202
 QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 Db 203 TGGTTTGACACTGGACTCTCAGATGCGCTGGAGAGGACAGTCTCTGTCTCACTG 262
 QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
 Db 263 GATGTGAAGACACTTCTCCCGACAGAACTCAAAAGTTAAGGTGTTGGAGATGTGATTGAG 322
 QY 96 IleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
 Db 323 GTGCATGAAGAAACATGAAGAGCCGAGGATGAACATGTTTCATCTCAGGAGGTTCAC 382
 QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
 Db 383 AGGAATACCGGATCCCGAGCTGATGATGACCTCTCACCATTACTTTCATCTCTCATCT 442
 QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
 Db 443 GATGGGGTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCT 490

QY 156 GluArgAlaIleProValSerArgGluGlyPro-----ThrSerAlaPro 171
 Db 491 GAGCGCACCATTCCTACCATCCCGTGAAGAGAGCTGTGTCTCACCGCAGCCCC 544
 RESULT 40
 BC007008
 LOCUS Homo sapiens crystallin, alpha B, mRNA (cdna clone MGC:12326 IMAGE:3933748), complete cds.
 DEFINITION BC007008
 ACCESSION BC007008
 VERSION BC007008.1 GI:13937812
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 744)
 AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,K.J., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,R.D., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 12477932
 REFERENCE 2 (bases 1 to 744)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: http://www-shgc.stanford.edu
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAL Plate: 16 Row: j Column: 15
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503056.
 Location/Qualifiers
 1..744
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:12326 IMAGE:3933748"
 /tissue_type="Heart"
 /clone_lib="NIH_MGC_74"
 /lab_host="DH10B"

```
/note="Vector: pDNR-LIB"
1. .744
/ gene="CRYAB"
/ note="synonyms: CRYA2, CTPP2, HSPB5"
/ db_xref="LocustID:1410"
/ db_xref="MIM:123590"
47_574
/ gene="CRYAB"
/ codon_start=1
/ product="crystallin, alpha B"
/ protein_id="AAH07008.1"
/ db_xref="GI:13937813"
/ db_xref="LocustID:1410"
/ db_xref="MIM:123590"
/ translation="MDIAHPHWRPPFPSPHPSRLPDQFFGHEHLESDLPFTSTSL
SPFYLRPSPFLRPSWFDTGLEMRLEKDRFVNLGVKHFPEELKVKVLGDVIEVHG
KHEERODEHGFISREFHRYKRI PADVDPLTITSSLSGDLTVNGPRKQVSGPERTIP
ITREEKPAVTAAPK"

ORIGIN
Alignment Scores:
Pred. No.: 1.96e-45 Length: 744
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 9 Gaps: 6

US-10-657-740-1 (1-173) x BC007008 (1-744)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGluProPheTyr---Pro 19
Db 47 ATGGACATCGCATCCACCCCTGGATCGCGCCCTCTTTCTTCCCTTCCACTCCCCC 106

Qy 20 SerArgLeuPheAspGlnPheGluGluLeuPheGluTyrAspLeuLeuProPhe 39
Db 107 AGCGGCTCTTTGACCACTTCTTGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 163

Qy 40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 164 ACGTCTACTCTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTTCTCGGGCACCACG 223

Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 224 TGGTTTGACACTGGACTCTCAGAGATCGCCTGGAAAGAGACAGGTTCTCTGTCAACCTG 283

Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 284 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGGAGATGTGATTGAG 343

Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 344 GTGCATGGAACAACATGAAGAGCGCAGATGACATGATGTTTCATCTCCAGGAGTTCCAC 403

Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 404 AGAAATACCGATCCCGATCGATGATGATGATGATGATGATGATGATGATGATGATGAT 463

Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 464 GATGGGTCTCTCACTGTAATGGACCAAGAAACAG-----GTCTCTGGCCCT 511

Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 512 GAGCGCACCATTCATCCATCCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 565

RESULT 41
AX888028
LOCUS AX888028 856 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 3891 from Patent EP1033401.
ACCESSION AX888028
VERSION AX888028.1 GI:40046779
KEYWORDS
```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Dumas Milne Edwards, J.B., Duclert, A. and Giordano, J.Y.
TITLE Expressed sequence tags and encoded human proteins
JOURNAL Patent: EP 1033401-A 3891 06-SEP-2000;
Genset (PR)
FEATURES
source
1. .856
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
4. .756
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE93953.1"
/db_xref="GI:40046780"
/translation="MTSAFKLDLFLPDMVXGRLLVDPDRINGTAKNGKALDXXQDPDP
DAIKMFVGZIPRSWSXKELTSLQTPHPTLADMDIAIHPWIRRPFPFPHSPSLPDQF
FGEHLESDLPFTSTSLSPFYLRPSPFLRPSWFDTGLEMRLEKDRFVNLGVKHF
PEELKVKVLGDVIEVHGKHEERODEHGFISREFHRYKRI PADVDPLTITSSLSGDL
TVNGPRKQVSGPERTIPTREEKPAVTAAPK"

ORIGIN
Alignment Scores:
Pred. No.: 2.35e-45 Length: 856
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x AX888028 (1-856)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGluProPheTyr---Pro 19
Db 229 ATGGACATCGCATCCACCCCTGGATCGCGCCCTCTTTCTTCCCTTCCACTCCCCC 288

Qy 20 SerArgLeuPheAspGlnPheGluGluLeuPheGluTyrAspLeuLeuProPhe 39
Db 289 AGCGGCTCTTTGACCACTTCTTGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 345

Qy 40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 346 ACGTCTACTCTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTTCTCGGGCACCACG 405

Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 406 TGGTTTGACACTGGACTCTCAGAGATCGCCTGGAGAGAGACAGGTTCTCTGTCAACCTG 465

Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 466 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTGGGAGATGTGATTGAG 525

Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 526 GTGCATGGAACAACATGAAGAGCGCAGATGACATGATGTTTCATCTCCAGGAGTTCCAC 585

Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 586 AGAAATACCGATCCCGATCGATGATGATGATGATGATGATGATGATGATGATGATGAT 645

Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 646 GATGGGTCTCTCACTGTAATGGACCAAGAAACAG-----GTCTCTGGCCCT 693

Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 694 GAGCGCACCATTCATCCATCCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747

RESULT 42

```
BD027638
LOCUS BD027638 856 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD027638
VERSION BD027638.1 GI:22569380
KEYWORDS JP 2001269182-A/3884.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
TITLE Sequence tag and encoded human protein.
JOURNAL Patent: JP 2001269182-A 3884 02-OCT-2001;
GENSET
COMMENT
OS Homo sapiens (human)
PN JP 2001269182-A/3884
PD 02-OCT-2001
PR 24-FEB-2000 JP 2000118773
PI JEAN BAPTIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
PI JORDAN
PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21,PC
PC C12N5/10,
PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00,PC
G06F15/40
CC
FH Key Location/Qualifiers
FT CDS 4..753.
FEATURES
source
1..856 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 2,356-45 Length: 856
Pred. No.: 489.00 Matches: 97
Score: 73.60% Conservative: 34
Percent Similarity: 54.49% Mismatches: 35
Best Local Similarity: 53.38% Indels: 12
Query Match: 6 Gaps: 6
DB:
US-10-657-740-1 (1-173) x BD027638 (1-856)
Qy 1 MetAspValThrIleSerProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 229 ATGGACATCGCATCCACACCCCTGGATCGCGCCCTTTCTTTCTTCCACTCCGCC 288
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
Db 289 AGCGCCCTCTTTGACAGTCTTCTCGAGAGCACCTGTTGGAGTCTGTATCTTTCCCG--- 345
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 346 ACGTCTACTTCTCCCTGAGTCCCTTTACCTTCGCCACCCTCTTCTCGCGGACCCAGC 405
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerArgAspArgAspLysPheValIlePheLeu 75
Db 406 TGGTTTGACACTGGACTCTCAGAGATCGCTGGAGAGAGACAGGTTCTCTGTCAACCTG 465
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 466 GATGTGAAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGGAGATGTGATTGAG 525
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 526 GTGCATGAAGCAATCATGAAGAGCGCCAGATGATGTTTTCATCTCCAGGAGATTCCAC 585
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 586 AGGAATACCGGATCCCGACTGATGTAGACCCCTCTCACCATTTACTTTCATCTCTGTCACT 645
```

```
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 646 GATGGGTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCT 693
Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 694 GAGCGCACCATTCCTCCATCACCCTGAAGAGAGCGCTGTGTGCACGCGACGCC 747
RESULT 43
AX899079
LOCUS AX899079 893 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 14942 from Patent EP1033401.
ACCESSION AX899079
VERSION AX899079.1 GI:40053992
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
TITLE Expressed sequence tags and encoded human proteins
JOURNAL Patent: EP 1033401-A 14942 06-SEP-2000;
GENSET
FEATURES
source
1..893 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores: 2,486-45 Length: 893
Pred. No.: 489.00 Matches: 97
Score: 73.60% Conservative: 34
Percent Similarity: 54.49% Mismatches: 35
Best Local Similarity: 53.38% Indels: 12
Query Match: 6 Gaps: 6
DB:
US-10-657-740-1 (1-173) x AX899079 (1-893)
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 266 ATGGACATCGCATCCACACCCCTGGATCGCGCCCTTTCTTTCTTCCACTCCGCC 325
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
Db 326 AGCGCCCTCTTTGACAGTCTTCTCGAGAGCACCTGTTGGAGTCTGTATCTTTCCCG--- 382
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 383 ACGTCTACTTCTCCCTGAGTCCCTTTACCTTCGCCACCCTCTTCTCGCGGACCCAGC 442
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerArgAspArgAspLysPheValIlePheLeu 75
Db 443 TGGTTTGACACTGGACTCTCAGAGATCGCTGGAGAGAGACAGGTTCTCTGTCAACCTG 502
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 503 GATGTGAAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGGAGATGTGATTGAG 562
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 563 GTGCATGAAGCAATCATGAAGAGCGCCAGATGATGTTTTCATCTCCAGGAGATTCCAC 622
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 623 AGGAATACCGGATCCCGACTGATGTAGACCCCTCTCACCATTTACTTTCATCTCTGTCACT 682
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 683 GATGGGTCTCTCACTGTGAATGAGCAAGAAACAG-----GTCTCTGGCCCT 730
```

```

QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 731 GAGCGACCATTCCTCCATCAGCGGTGAAGAGAGCTGTGTCTACCGGACCCCC 784

RESULT 44
BD034612
LOCUS BD034612 893 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD034612
VERSION BD034612.1 GI:22576354
KEYWORDS JP 2001269182-A/10858.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.
TITLE Sequence tag and encoded human protein
JOURNAL Patent: JP 2001269182-A 10858 02-OCT-2001;
GENSET
COMMENT
OS Homo sapiens (human)
PN JP 2001269182-A/10858
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US -60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC
FH key Location/Qualifiers.
FEATURES
source
1..893 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 2,486-45 Length: 893
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x BD034612 (1-893)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 266 ATGGACATCGGCATCCACCCCTGGATCCCGCCCTTCTTCTTCCATCCCTCC 325

QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 326 AGCCGCTCTTTGACCAAGTTCTTCGGAGAGACCTGTTGGAGTCTGATCTTTCC 382

QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 383 AGCTTACTTCCCTGAGTCCCTTCTACCTTCGCCACCCCTCTCTCTCGGGACCC 442

QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 443 TGGTTTGACACTGGACTCTCAGAGATCGCCCTGGAGAGAGACAGGTTCTCTGTCA 502

QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 503 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGTTAGTTGGAGATGTGATTG 562

QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 96 ILEHISGLYLYSHISASNGLUARGGLNASPASPHisGLYTYRILESERARGGLUPHEHIS
:::|||||

```

```

Db 563 GTGCATGGAAACATGAGAGCGCCAGCATGAACATGGTTTCATCTCCAGGAGTTCCAC 622
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 623 AGGAAATACCGGATCCAGCTGATGATAGACCTCTCAGCATTTACTTCTATCCCTGTCT 682

QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 683 GATGGGGTCTTCATCTGTAATGGACCAAGGAACAG-----GTCTCTGGCCCT 730

QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 731 GAGCGACCATTCCTCCATCAGCGGTGAAGAGAGCTGTGTCTACCGGACCCCC 784

RESULT 45
LOCUS AX899075 911 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 14938 from Patent EP1033401.
ACCESSION AX899075
VERSION AX899075.1 GI:40053988
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Dumas Milne Edwards, J.B., Duclert, A. and Giordano, J.Y.
TITLE Expressed sequence tags and encoded human proteins
JOURNAL Patent: EP 1033401-A 14938 06-SEP-2000;
GENSET
FEATURES
source
1..911 Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 2,546-45 Length: 911
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x AX899075 (1-911)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 284 ATGGACATCGGCATCCACCCCTGGATCCCGCCCTTCTTCTTCCATCCCTCC 343

QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 344 AGCCGCTCTTTGACCAAGTTCTTCGGAGAGACCTGTTGGAGTCTGATCTTTCC 400

QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 401 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCCACCCCTCTCTCTCGGGACCC 460

QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 461 TGGTTTGACACTGGACTCTCAGAGATCGCCCTGGAGAGAGACAGGTTCTCTGTCA 520

QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 521 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGTTAGTTGGAGATGTGATTG 580

QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 581 GTGCATGGAAACATGAGAGCGCCAGCATGAACATGGTTTCATCTCCAGGAGTTCCAC 640

QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 116 ARGARGTYRARGLEUPROSERASNVALASPGLNSERALALESERCYSSERLEUSERALA
:::|||||

```

```

Db      641 AGGAAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTTCTATCTCCCTGTCATCT 700
Qy      136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db      701 GATGGGGTCTCCTCACTGTGAATGAGCAAGAAACAG-----GTCTCTGGCCCT 748
Qy      156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db      749 GAGCGCACCATTTCCCATCACCGTGAAGAGAGCTGTGTGTCACCGAGCCCCC 802

RESULT 46
LOCUS   BD034608
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD034608
VERSION   BD034608.1 GI:22576350
KEYWORDS JP 2001269182-A/10854.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 911)
AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
TITLE      Sequence tag and encoded human protein
JOURNAL    Patent: JP 2001269182-A 10854 02-OCT-2001;
          GENSET
COMMENT   OS Homo sapiens (human)
          PN JP 2001269182-A/10854
          PD 02-OCT-2001
          PF 24-FEB-2000 JP 2000118773
          PR 26-FEB-1999 US 60/122487
          PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
          FI JORDAN
          PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
          C12N5/10,
          PC C12P21/02, C12P21/08, C12Q1/68 //G06F17/30, C12N15/00, C12N5/00, PC
          G06F15/40
          CC
          FH Key Location/Qualifiers.
FEATURES
          source
          1..911
          /organism="Homo sapiens"
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 2,546-45 Length: 911
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x BD034608 (1-911)
Qy      1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      284 ATGGACATCGCATCCACACCCCTGGATCGCCGCCCTTCTTTCTTCTCACTCCCCC 343
Qy      20 SerArgLeuPheAspGluPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db      344 AGCCGCCCTTTTGACAGTTCTTCGGAGAGACACCTGTGGAGTCTGATCTTTTCCCG--- 400
Qy      40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db      401 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCCACCCCTCTCTCTCGGGCACCACCG 460
Qy      56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db      461 TGGTTTGACACTGGGACTCTCAGAGATGCGCCTGGAGAGAGCAGGTTCTCTGTCAACCTG 520
Qy      76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95

```

```

Db      521 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTTGGGAGATGTGATTGAG 580
Qy      96 IleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115
Db      581 GTGCATGAAACATGATGAAGAGCGCAGATGAACATGGTTTTCATCTCCAGAGGAGTTCCAC 640
Qy      116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db      641 AGGAAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTTCTATCTCTGTCATCT 700
Qy      136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db      701 GATGGGGTCTCCTCACTGTGAATGAGCAAGAAACAG-----GTCTCTGGCCCT 748
Qy      156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db      749 GAGCGCACCATTTCCCATCACCGTGAAGAGAGCTGTGTGTCACCGAGCCCCC 802

RESULT 47
LOCUS   AX899073
DEFINITION Sequence 14936 from Patent EP1033401.
ACCESSION AX899073
VERSION   AX899073.1 GI:40053986
KEYWORDS Homo sapiens (human)
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
TITLE      Expressed sequence tags and encoded human proteins
JOURNAL    Patent: EP 1033401-A 14936 06-SEP-2000;
          GENSET
FEATURES
          source
          1..927
          /organism="Homo sapiens"
          /mol_type="unassigned DNA"
          /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 2,6e-45 Length: 927
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x AX899073 (1-927)
Qy      1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      300 ATGGACATCGCATCCACACCCCTGGATCGCCGCCCTTCTTTCTTCTTCCACTCCCCC 359
Qy      20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db      360 AGCCGCCCTTTTGACAGTTCTTCGGAGAGACACCTGTGGAGTCTGATCTTTTCCCG--- 416
Qy      40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db      417 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCCACCCCTCTCTCTCGGGCACCACCG 476
Qy      56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db      477 TGGTTTGACACTGGGACTCTCAGAGATGCGCCTGGAGAGAGCAGGTTCTCTGTCAACCTG 536
Qy      76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
Db      537 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGTTGGGAGATGTGATTGAG 596
Qy      96 IleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115

```

```

Db      597 GTGCGATGGAAACATGAAGAGCGCCAGGATGAACATGTTTCATCTCCAGGAGTTCCAC 656
Qy      116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db      657 AGGAATACCGATCCCGATGATGATAGACCTCTCACCATTACTTTCATCTCTGTCATCT 716
Qy      136 AspGlyMetLeuThrPheCysGlyProlysisGlnThrGlyLeuAspAlaThrHisAla 155
Db      717 GATGGGTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGCCCT 764
Qy      156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db      765 GAGCGCACCATTCCTCCATCACCCTGAGAGAGACCTGTGTCTCAGCGAGCCGCC 818

RESULT 48
LOCUS   BD034606                      927 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD034606
VERSION   BD034606.1 GI:22576348
KEYWORDS JP 2001269182-A/10852.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 927)
AUTHORS   Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
TITLE      Sequence tag and encoded human protein
JOURNAL    Patent: JP 2001269182-A 10852 02-OCT-2001;
          GENSET
COMMENT   OS Homo sapiens (human)
          PN JP 2001269182-A/10852
          PD 02-OCT-2001
          PE 24-FEB-2000 JP 2000118773
          PR 28-FEB-1999 US 60/122487
          PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
          PJ JORDAN
          PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21, PC
          C12N5/10,
          PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00, PC
          G06F15/40
          CC
          FH Key Location/Qualifiers.
FEATURES             source
          1..927
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
ORIGIN
Alignment Scores:      2.6e-45      Length:      927
Pred. No.:             489.00      Matches:      97
Score:                 73.60%      Conservative: 34
Percent Similarity:    54.43%      Mismatches:   35
Best Local Similarity: 53.38%      Indels:       12
Query Match:           6           Gaps:         6
DB:

US-10-657-740-1 (1-173) x BD034606 (1-927)

Qy      1 MetAepValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      300 ATGGACATCGCATCCACCACCCCTGGATCCGCGCCCTTCTTCTTCTTCCATCCCC 359
Qy      20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
Db      360 AGCCGCCCTCTTGACAGTCTTTCGAGAGACCTGTGGAGTCTGTGATCTTTTCCG--- 416
Qy      40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db      417 ACGTCTACTTCCCTGAGTCCCTTCTTACCTTCGGCCACCCTCTTCTTCTCGGGGACCC 476

```

```

Qy      56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db      477 TGGTTTGACATGGACTCTCAGATCGCTCGAGAGAGGACAGGTTCTCTGTCAACCTG 536
Qy      76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
Db      537 GATGTGAAGCACTTCTCCAGAGAACTCAAGTTAAGGTGTTGGGAGATGTGATTGAG 596
Qy      96 IleHisGlyLysHisAsnGlnArgGlnAspHisGlyTyrIleSerArgGluPheHis 115
Db      597 GTGCATGAAACATCAGAGAGCGCCAGCATGAACATGTTTCATCTCCAGGAGTTCCAC 656
Qy      116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db      657 AGGAATACCGATCCCGATGATGATAGACCTCTCACCATTACTTTCATCTCTGTCATCT 716
Qy      136 AspGlyMetLeuThrPheCysGlyProlysisGlnThrGlyLeuAspAlaThrHisAla 155
Db      717 GATGGGTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGCCCT 764
Qy      156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db      765 GAGCGCACCATTCCTCCATCACCCTGAGAGAGACCTGTGTCTCAGCGAGCCGCC 818

RESULT 49
LOCUS   AX888027                      942 bp      DNA      linear      PAT 18-DEC-2003
DEFINITION Sequence 3890 from Patent EP1033401.
ACCESSION AX888027
VERSION   AX888027.1 GI:40046777
KEYWORDS  Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS   Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
TITLE      Expressed sequence tags and encoded human proteins
JOURNAL    Patent: EP 1033401-A 3890 06-SEP-2000;
          Genset (FR)
FEATURES             source
          1..942
             /organism="Homo sapiens"
             /mol_type="unassigned DNA"
             /db_xref="taxon:9606"
             315..842
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="CAE93952.1"
             /db_xref="GI:40046778"
             /translation="MDIAIHPIWIRRRPFPFPHSRPLFDQFFGBHLLESDLPPTSTSL
             SPFLRPSPFLRAPSWFDTLSEMRLEKDRSVNLVDVKHFSPEELKVKVLGVDIVHVG
             KHEERQDEHGFISEFHRKYRIPADVDPLTITSSLSGDLTVNGPRKQVSGPERTIP
             ITREKPAVTAAPKK"
ORIGIN
Alignment Scores:      2.65e-45      Length:      942
Pred. No.:             489.00      Matches:      97
Score:                 73.60%      Conservative: 34
Percent Similarity:    54.49%      Mismatches:   35
Best Local Similarity: 53.38%      Indels:       12
Query Match:           6           Gaps:         6
DB:

US-10-657-740-1 (1-173) x AX888027 (1-942)

Qy      1 MetAepValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db      315 ATGGACATCGCATCCACCACCCCTGGATCCGCGCCCTTCTTCTTCTTCCATCCCC 374
Qy      20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
Db      375 AGCCGCCCTCTTGACAGTCTTCTTACCTTCGGAGAGACCTGTGGAGTCTGTGATCTTTTCCCG--- 431

```

Qy	40	LeuSerSerThrIleSerProTyrTyArgGln-----SerLeuPheArg-----Thr	55
Db	432	ACGTCCTACTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTCTCTCGCGGCACCCGAC	491
Qy	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgaspArgLysPheValIlePheLeu	75
Db	492	TGGTTTGACACTGGACACTCTCAGAGATGCGCTTGAGAGAGGACAGGTTCTCTGTCAACCTG	551
Qy	76	AspVallylHisPheSerProGluAspLeuThrVallylsValGlnAspAspPheValGlu	95
Db	552	GATCTGAAGCATTCTCCCCAGAGGAAGTCAAAAGTTAAGGTGTTGGGAGATGTGATTGAG	611
Qy	96	IleHisGlyLysHisAenGluuArgGlnAspAspHisGlyTyrIleSerArgGluPheHis	115
Db	612	GTGCATGGAACATGAAGAGCGCAGGATGAACATGTTTTTCATCTCCAGGAGATTCCAC	671
Qy	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla	135
Db	672	AGGAAATACCGGATCCCGAGCTGTGTGAGACCCCTCTCACATTACTTCATCCCTGTGATCT	731
Qy	136	AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla	155
Db	732	GATGGGTCTCATGTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCT	779
Qy	156	GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro	171
Db	780	GAGCGCACCATTTCCCATCATCCCGTGAAGAGAAGGCTGCTGTGCACCGGACGCCCC	833

RESULT 50

BD027637	BD027637	942 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	Sequence tag and encoded human protein.				
DEFINITION	BD027637				
ACCESSION	BD027637.1	GI:22569379			
VERSION	JP 2001269182-A/3883				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 942)				
AUTHORS	Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.				
TITLE	Sequence tag and encoded human protein				
JOURNAL	Patent: JP 2001269182-A 3883 02-OCT-2001; GENSET				

ORIGIN

Alignment Scores:

Alignment Scores:		
Pred. No.:	2.65e-45	Length: 942
Score:	489.00	Matches: 97
Percent Similarity:	73.60%	Conservative: 34
Best Local Similarity:	54.49%	Mismatches: 35
Query Match:	53.38%	Indels: 12
DB:	6	Gaps: 6

```

US-10-657-740-1 (1-173) x BD027637 (1-942)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 315 ATGGACATCGCATCCACACCCCTGGATCGCGCGCCCTCTCTTCCATTCCACTCCCCC 374

Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 375 AGCGCGCTCTTTGACCACTAGTTCTTCGGAGAGCACCTGTTGGAGTCTGATCTTTTCCG-- 431

Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 432 ACGTCTACTTCCCTGAGTGCCCTCTACTCTCGGCCACCTCTCTTCGCGGCACCCAGC 491

Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 492 TGGTTTGACACTGGACACTCTAGAGATGCGCTCGGAGAGGACAGGTTCCTGTCAACCTG 551

Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 552 GATGTGAAGCACTTCTCCCGAGAGGAATCAAAGTTAAGGTGTGGGAGATGTGATTCAG 611

Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 612 GTGCATCGAAAAACATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGAGTTCAC 671

Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 672 AGGAAATACCGGATCCCGACTGATGTAGACCTCTCACCATTTACTTCATCCCTGTGATCT 731

Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 732 GATGGGTCTCATGTGATGGACCAAGGAACAG-----GTCTCTGGCCCT 779

Qy 156 GluArgAlaIleProValSerArgGluLulysPro-----ThrSerAlaPro 171
Db 780 GAGCGCACCATTTCCCATCACCCTGTGAAGAGAAGCTGCTGTCAACCGAGCCCCC 833

```

Search completed: May 30, 2005, 07:14:52
Job time : 4226.45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 29, 2005, 22:54:38 ; Search time 526.598 Seconds
(without alignments)
1944.776 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVTIQHFWKRTLGPFYPS.....HAERAIPVSREKETSAPSS 173

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/usPTO_epool/US10657740/runat_27052005_165252_3305/app_query.fasta_1.590
-DB=N_Geneseq_16Dec04 -QMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOPECL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bioSum62 -TRANS=human40.cdi
-LIST=150 -DOCLFMT=200 -THR SCORES=ptc -THR MAX=100 -THR MTN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10657740 @CGN 1 1.796 @runat_27052005_165252_3305 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*

1: geneseq1980s:*

2: geneseq1990s:*

3: geneseq2000s:*

4: geneseq2001as:*

5: geneseq2001bs:*

6: geneseq2002as:*

7: geneseq2002bs:*

8: geneseq2003as:*

9: geneseq2003bs:*

10: geneseq2003cs:*

11: geneseq2003ds:*

12: geneseq2004as:*

13: geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	871	95.1	1056	10 ADB52521	Adb52521 Primary r
2	871	95.1	1056	10 ADF30549	Adf30549 Rat angio
3	849.5	92.7	1271	10 ADF30547	Adf30547 Rat angio
4	702	76.6	531	8 ABX12062	Abx12062 Human alp
5	639	69.8	372	13 ADQ78288	Adq78288 Human N-t

6	493	53.8	548	6	ABZ35138	Abz35138 Human gen
7	489	53.4	528	13	ADR46388	Adr46388 Human len
8	489	53.4	537	10	ADE75374	Ade75374 Human mut
9	489	53.4	537	10	ADE75375	Ade75375 Human wil
10	489	53.4	691	2	AAZ39668	Aax39668 Renal can
11	489	53.4	691	6	ABL65204	Ab165204 Lung can
12	489	53.4	691	6	ABL62427	Ab162427 Colon ade
13	489	53.4	691	6	ABT10966	Abt10966 Human bre
14	489	53.4	691	6	ABK64671	Abk64671 Human ben
15	489	53.4	691	10	ADF09686	Adf09686 Human cry
16	489	53.4	691	12	ADN05800	Adn05800 Antipsori
17	489	53.4	691	13	ADR24825	Adr24825 Breast ca
18	489	53.4	856	3	AAZ39668	Aax39668 Renal can
19	489	53.4	893	3	AAZ39668	Aax39668 Renal can
20	489	53.4	911	3	AAZ39668	Aax39668 Renal can
21	489	53.4	927	3	AAZ39668	Aax39668 Renal can
22	489	53.4	942	3	AAZ39668	Aax39668 Renal can
23	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
24	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
25	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
26	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
27	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
28	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
29	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
30	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
31	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
32	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
33	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
34	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
35	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
36	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
37	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
38	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
39	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
40	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
41	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
42	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
43	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
44	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
45	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
46	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
47	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
48	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
49	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
50	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
51	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
52	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
53	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
54	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
55	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
56	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
57	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
58	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
59	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
60	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
61	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
62	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
63	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
64	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
65	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
66	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
67	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
68	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
69	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
70	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
71	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
72	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
73	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
74	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
75	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
76	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
77	489	53.4	1036	2	AAZ33574	Aax33574 Human bre
78	489	53.4	1036	2	AAZ33574	Aax33574 Human bre

79	283	30.9	1379	5	AAD02211	Partial p
80	276	30.1	503	6	ABL99507	Target ca
81	275	30.0	558	10	ADK12085	Breast ca
82	266.5	29.1	604	6	ABK63074	Rat sequ
83	266.5	29.1	604	12	ADP72259	Renal tox
84	266.5	29.1	2874	4	ABL07842	Drosophil
85	266.5	29.1	3119	4	ABL09418	Drosophil
86	266	29.0	410	6	AB570907	Deer cdna
87	265.5	29.0	909	2	AQ03969	Clone P25
88	263.5	28.8	722	13	ADT6608	Murine he
89	259.5	28.3	654	5	AA542494	Human cdn
90	257	28.1	1700	4	AA563378	Human cdn
91	256.5	28.0	755	2	AQ03970	Clone P25
92	251	27.4	376	4	AA119387	Human bre
93	251	27.4	549	3	AAC93846	Cat flea
94	247	27.0	550	12	ADL11752	Cat flea
95	241.5	26.4	708	5	ADL40880	Human ova
96	234.5	25.6	614	5	ADI75644	Human ova
97	234.5	25.6	614	5	ADI69298	Human ova
98	233.5	25.5	954	5	ADL62061	Human ova
99	225.5	24.6	542	12	ADL11704	Cat flea
100	223.5	24.4	500	9	ACH34922	Human end
101	223	24.3	467	6	ABV94661	Human pan
102	220.5	24.1	482	9	ACH17221	Human adu
103	215	23.5	870	4	ABL07721	Drosophil
104	215	23.5	2870	4	ABL07720	Drosophil
105	212	23.1	49634	6	ABL68647	Kidney ca
106	205.5	22.4	212	8	ABT33639	Anticance
107	202.5	22.1	937	4	ABL07315	Drosophil
108	202.5	22.1	2651	4	ABL07714	Drosophil
109	202.5	22.1	2937	4	ABL07314	Drosophil
110	198.5	21.7	396	6	ABL63575	Breast ca
111	198.5	21.7	396	6	ABL63995	Breast ca
112	196	21.4	1206	4	ABL07725	Drosophil
113	196	21.4	3206	4	ABL07724	Drosophil
114	193	21.1	600	4	ABL07321	Drosophil
115	193	21.1	2600	4	ABL07320	Drosophil
116	192	21.0	6806	9	ACF25379	Rat alpha
117	191.5	20.9	445	9	ACH23297	Human adu
118	188	20.5	4206	6	ABZ35176	Human gen
119	188	20.5	4206	10	ABE84874	Farnesyl
120	187	20.4	2632	8	ABX63312	Human cdn
121	182.5	19.9	393	3	AAC06428	Human sec
122	182.5	19.9	487	13	ADS54436	Bacterial
123	182.5	19.9	487	13	ADS54437	Bacterial
124	180.5	19.7	565	11	ACN79216	Breast ca
125	179	19.5	218	12	AA109000	Human bre
126	179	19.5	1811	12	ADJ75836	Marker ge
127	174.5	19.1	1367	4	ABL07295	Drosophil
128	174.5	19.1	3367	4	ABL07294	Drosophil
129	174	19.0	510	6	AQ14062	Oligonuc
130	174	19.0	510	6	AQ14063	Oligonuc
131	172.5	18.8	1495	10	ADI22459	Rat liver
132	171.5	18.7	1627	2	AAZ21959	CDNA enco
133	171.5	18.7	1627	9	ACH04117	Human cdn
134	171.5	18.7	1843	4	AAE74328	Human H11
135	171.5	18.7	1843	4	AAE74307	Human H11
136	171.5	18.7	1854	5	ABX71326	Human met
137	171.5	18.7	2004	5	ADM19232	Novel hum
138	171.5	18.7	2010	12	ADJ74941	Marker ge
139	171.5	18.7	2010	13	ADR25621	Breast ca
140	171.5	18.7	2012	4	AAK52551	Human pol
141	171.5	18.7	2048	4	AAK51567	Human pol
142	165	18.0	651	4	ABL07715	Drosophil
143	165	18.0	2651	4	ABL07714	Drosophil
144	165	18.0	2937	4	ABL07314	Drosophil
145	163.5	17.8	2036	10	ADI22637	Human liv
146	163	17.8	509	3	AAAC01050	Human sec
147	163	17.8	546	3	AAAC01053	Human sec
148	163	17.8	564	3	AAAC01049	Human sec
149	163	17.8	580	3	AAAC01051	Human sec
150	163	17.8	595	3	AAAC01048	Human sec

ALIGNMENTS

RESULT 1

ADBS2521

ID ADBS2521 standard; DNA; 1056 BP.

XX

AC ADBS2521;

XX

DT 04-DEC-2003 (first entry)

XX

XX Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3063.

XX

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;

KW toxicity marker; toxicity progression; drug screening;

KW primary rat hepatocyte toxicity modelling; gene; ds.

XX

OS Rattus norvegicus.

XX

PN WO2003065993-A2.

XX

PD 14-AUG-2003.

XX

XX 04-FEB-2003; 2003WO-US003482.

XX

XX 04-FEB-2002; 2002US-0353171P.

XX

XX 13-MAR-2002; 2002US-0363534P.

XX

XX 08-APR-2002; 2002US-0370248P.

XX

XX 10-APR-2002; 2002US-0371134P.

XX

XX 10-APR-2002; 2002US-0371135P.

XX

XX 10-APR-2002; 2002US-0371150P.

XX

XX 11-APR-2002; 2002US-0371413P.

XX

XX 19-APR-2002; 2002US-0373601P.

XX

XX 19-APR-2002; 2002US-0373602P.

XX

XX 22-APR-2002; 2002US-0374139P.

XX

XX 08-MAY-2002; 2002US-0378370P.

XX

XX 09-MAY-2002; 2002US-0378652P.

XX

XX 09-MAY-2002; 2002US-0378653P.

XX

XX 09-MAY-2002; 2002US-0378655P.

XX

XX 09-JUL-2002; 2002US-0394230P.

XX

XX 09-JUL-2002; 2002US-0394253P.

XX

XX 04-SEP-2002; 2002US-0407688P.

XX

XX 28-JAN-2003; 2003US-0442900P.

XX

XX (GENE-) GENE LOGIC INC.

XX

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX

XX Elashoff M;

XX

XX WPI; 2003-731472/69.

XX

XX Determining if a compound induces a toxic effect on a tissue or cell, for

XX

XX identifying hepatotoxic compounds, comprises comparing a gene expression

XX

XX profile of a tissue or cell sample to a database of Tox mean and non-Tox

XX

XX mean values.

XX

XX Claim 44; SEQ ID NO 3063; 874pp; English.

XX

XX The present invention describes a method for determining whether a

XX

XX compound induces a toxic effect on a tissue or cell. The method comprises

XX

XX preparing a gene expression profile of a tissue or cell sample exposed to

XX

XX the compound, and comparing the gene expression profile to a database

XX

XX comprising data or information on the Tox mean and non-Tox mean value.

XX

XX The method is useful for predicting or identifying at least one toxic

XX

XX effect, particularly hepatotoxicity, of a test or unknown compound. The

XX

XX genes listed in the specification are useful as diagnostic or toxicity

XX

XX markers for the prediction or identification of the physiological state

XX

XX of tissue or cell sample that has been exposed to a compound, or to

XX

XX identify or predict the toxic effects of a compound or an agent. These

XX

XX may also be used as markers for monitoring toxicity progression or for

XX

XX drug screening. The present sequence represents a primary rat hepatocyte

XX

XX toxicity modelling related gene sequence from the present invention.

XX

SQ Sequence 1056 BP; 199 A; 333 C; 269 G; 255 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,07e-102 Length: 1056
Score: 871.00 Matches: 164
Percent Similarity: 97.69% Conservative: 5
Best Local Similarity: 94.80% Mismatches: 4
Query Match: 95.09% Indels: 0
DB: 10 Gaps: 0

US-10-657-740-1 (1-173) x ADB52521 (1-1056)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
DB 13 ATGGAGCTCACCATCCAGCACCCCTTGGTTCAAGCGCGCCCTTGGGCGCCTTCTACCCGACG 72
QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
DB 73 CGACTGTTCCAGCAGTCTTGGCGAGGCGCTTTTGAATACAGCTGCTGCCCTTCTCTG 132
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
DB 133 TCTTCCACCATCAGCCCTTACTACCGCCAGTCTCTCTCCGCACACAGTGTGGACTCCGCG 192
QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
DB 193 ATCTCTGAGTCCGATCAGCCGGGACAAGTTTGTTCATCTTCTTGGATGGAAGCACTTC 252
QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValIleHisGlyLysHis 100
DB 253 TCTCTGAGGACTCACCCTGGAAGGTACTGGAAGATTCTGTGGAGATCCATGGCAACAC 312
QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
DB 313 AACGAGAGCAGATGACCATCGGTACATTTCCGTGAATTTCCACGTCGCTACCGTCTG 372
QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
DB 373 CCTTCCAAATGTGACACAGTCCGCGCTCTCTGCTGCTGCGGATGCGATGCTGACC 432
QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
DB 433 TTCTCTGGCCCCAAGTCTGAGTGTGGATGTGGCCACAGCGAGAGGGGCCATTCCC 492
QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
DB 493 GTGTACGGGAGGAGAGCCAGCTCGGCACCCCTCGTCC 531

RESULT 2

ID ADF30549

ADP30549 standard; cdNA; 1056 BP.

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX AC ADF30549;

XX (PROC) PROCTER & GAMBLE CO.

XX Peters KG, Thompson LJ, Wang F, Greis KD;

XX WPI; 2003-711557/67.

XX P-PSDB; ADF30550.

XX Treating angiogenesis-mediated disorder, e.g., retinal or choroidal

XX neovascularization or diseases associated with chronic inflammation,

XX myocardial ischemia, stroke, coronary artery disease or peripheral

XX vascular disease.

XX Disclosure; SEQ ID NO 112; 26pp; English.

XX The invention relates to a method of treating an angiogenesis-mediated

XX disorder in a subject. The method is useful for treating angiogenesis-

XX mediated disorder, e.g., retinal or choroidal neovascularisation or

XX diseases associated with chronic inflammation, myocardial ischaemia,

XX stroke, coronary artery disease or peripheral vascular disease. The

XX present sequence is used in the exemplification of the invention.

XX SQ Sequence 1056 BP; 199 A; 333 C; 269 G; 255 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6,07e-102 Length: 1056
Score: 871.00 Matches: 164
Percent Similarity: 97.69% Conservative: 5
Best Local Similarity: 94.80% Mismatches: 4
Query Match: 95.09% Indels: 0
DB: 10 Gaps: 0

US-10-657-740-1 (1-173) x ADF30549 (1-1056)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
DB 13 ATGGAGCTCACCATCCAGCACCCCTTGGTTCAAGCGCGCCCTTGGGCGCCTTCTACCCGACG 72
QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
DB 73 CGACTGTTCCAGCAGTCTTGGCGAGGCGCTTTTGAATACAGCTGCTGCCCTTCTCTG 132
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
DB 133 TCTTCCACCATCAGCCCTTACTACCGCCAGTCTCTCTCCGCACACAGTGTGGACTCCGCG 192
QY 61 IleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPhe 80
DB 193 ATCTCTGAGTCCGATCAGCCGGGACAAGTTTGTTCATCTTCTTGGATGGAAGCACTTC 252
QY 81 SerProGluAspLeuThrValLysValGlnAspAspPheValIleHisGlyLysHis 100
DB 253 TCTCTGAGGACTCACCCTGGAAGTACTGGAAGATTCTGTGAGATCCATGGCAACAC 312
QY 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
DB 313 AACGAGAGCAGATGACCATCGGTACATTTCCGTGAATTTCCACGTCGCTACCGTCTG 372
QY 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
DB 373 CCTTCCAAATGTGACACAGTCCGCGCTCTCTGCTGCTGCGGATGCGATGCTGACC 432
QY 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
DB 433 TTCTCTGGCCCCAAGTCTGAGTGTGGATGTGGCCACAGCGAGAGGGGCCATTCCC 492
QY 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
DB 493 GTGTACGGGAGGAGAGCCAGCTCGGCACCCCTCGTCC 531

RESULT 3

ADP30547

ID ADF30547 standard; cdNA; 1271 BP.

```

XX ADF30547;
AC
XX 12-FEB-2004 (first entry)
DT
XX Rat angiogenesis modulating protein cDNA #28.
DE
XX ss; gene; rat; angiogenesis; angiogenesis modulating protein;
KW retinal neovascularisation; choroidal neovascularisation;
KW chronic inflammation; myocardial ischaemia; stroke;
KW coronary artery disease; peripheral vascular disease.
XX
XX Rattus norvegicus.
OS
XX US2003162706-A1.
PN
XX 28-AUG-2003.
PD
XX 10-DEC-2002; 2002US-00316253.
PF
XX 08-FEB-2002; 2002US-035295P.
PR
XX 26-JUN-2002; 2002US-0391758P.
XX
XX (PROC ) PROCTER & GAMBLE CO.
PA
XX Peters KG, Thompson LJ, Wang F, Greis KD;
PI
XX WPI; 2003-711557/67.
XX
XX P-PSDB; ADF30548.
DR
XX
XX Treating angiogenesis-mediated disorder, e.g., retinal or choroidal
PT neovascularization or diseases associated with chronic inflammation,
PT myocardial ischemia, stroke, coronary artery disease or peripheral
PT vascular disease.
XX
XX Disclosure; SEQ ID NO 110; 26pp; English.
XX
XX The invention relates to a method of treating an angiogenesis-mediated
CC disorder in a subject. The method is useful for treating angiogenesis-
CC mediated disorder, e.g., retinal or choroidal neovascularisation or
CC diseases associated with chronic inflammation, myocardial ischaemia,
CC stroke, coronary artery disease or peripheral vascular disease. The
CC present sequence is used in the exemplification of the invention.
XX
XX SQ Sequence 1271 BP; 263 A; 384 C; 315 G; 309 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4.76e-99 Length: 1271
Score: 849.50 Matches: 164
Percent Similarity: 86.22% Conservative: 5
Best Local Similarity: 83.67% Mismatches: 4
Query Match: 92.74% Indels: 23
DB: 10 Gaps: 1

US-10-657-740-1 (1-173) x ADF30547 (1-1271)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
Db 159 ATGGAGCTCACCATCAGCACCCCTTGGTTCAAGCGCGCCCTGGGGCCCTTCTACCCGAGC 218
QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db 219 CGACTGTTCCAGCAGTCTTCGGCGAGGGCCCTTTTGAATACAGCTGCTGCTCCCTTCCG 278
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 279 TCTTCCACCATCAGCCCTTACTACCGCAGCTCTCTTCGGCAGCAGTGTGGACTCCGCG 338
QY 61 IleSerGlu----- 63
Db 339 ATCTCTGAGCTCATGACCCCATATGTGTTTGAATGCACCAACACCATGCTGGAAACCCC 398
QY 64 -----ValArgSerAspArgAspLysPheValIlePheLeuAspVal 77

```

```

Db 399 AAGAACACCCCGCAAGTCCGATCTCAGCGGACAAAGTTTGTTCATCTTCTGGATGG 458
QY 78 LysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHis 97
Db 459 AAGCACTTCTCTCTCTGAGGACCTCACCCTGAAGGTACTGGAGATTTCGTGGAGATCCAT 518
QY 98 GlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArg 117
Db 519 GCACAAACAACAGGAGGAGGAGGATGACCATGGCTACATTTCCCGTGAATTTCACCGTCGC 578
QY 118 TyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGly 137
Db 579 TACCGTCTGCGCTTCCAAATGTGACCACTCCGCGCTCTCTCTGCTCTGTCTCGGATGCG 638
QY 138 MetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArg 157
Db 639 ATGCTGAGCTTCTCTGCGGCCCAAGGTCCAGTCTGGCTTGGATGCTGGCCACAGCGAGAGG 698
QY 158 AlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 699 GCCATTCCCGTGTACGGGAGGAGGAGGCCAGCTCGGACCCCTCGTCC 746

RESULT 4
ABX12062
ID ABX12062 standard; DNA; 531 BP.
XX
AC ABX12062;
XX
DT 16-MAY-2003 (first entry)
XX
DE Human alpha BNAC crystallin chimera DNA.
XX
KW Human; ds; gene; chimera; alpha BNAC crystallin; protein shelf life;
KW protein aggregation; accessible hydrophobic region increase; mutant;
KW larger size oligomer formation; intersubunit interaction increase;
KW larger aggregate formation; larger porous oligomer formation;
KW increased ellipticity; less solvent accessible tryptophan;
KW increased chaperone-like activity; alpha A crystallin;
KW alpha B crystallin.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS
FT 1..531
FT /tag= a
FT /product= "Alpha BNAC crystallin"
FT /partial
FT /note= "No stop codon given"
FT misc_feature 1..247
FT /tag= b
FT /note= "Derived from DNA sequence encoding N-terminal
FT sequence of alpha B crystallin"
FT misc_feature 248..531
FT /tag= c
FT /note= "Derived from DNA sequence encoding N-terminal
FT sequence of alpha A crystallin"
XX
PN US2002177192-A1.
XX
PD 28-NOV-2002.
XX
PF 26-MAR-2002; 2002US-00105427.
XX
PR 28-MAR-2001; 2001US-0279223P.
XX (KUMA/) KUMAR L V S.
XX (RAOC/) RAO C M.
XX Kumar LVS, Rao CM;
XX WPI; 2003-298776/29.
DR

```

DR P-PSDB; ABG76084.
 XX New chimera alpha BNAC nucleic acid, useful for preventing aggregation of
 PT proteins and also for increasing shelf life of proteins of pharmaceutical
 PT value.
 XX Claim 1; Fig 9; 17pp; English.
 XX The invention relates to a chimera alpha BNAC polynucleotide that encodes
 CC a chimeric alpha BNAC polypeptide. The polypeptide is useful for
 CC preventing protein aggregation. The polypeptide is also useful for
 CC increasing the shelf life of proteins of pharmaceutical value. The
 CC polypeptide shows an increase in accessible hydrophobic regions, forms
 CC larger size oligomers, shows an increase in intersubunit interaction,
 CC forms larger aggregates, forms larger porous oligomers and shows
 CC increased ellipticity as compared to eye lens crystallins alpha A and
 CC alpha B. The tryptophan residues in the polypeptide are less solvent
 CC accessible as compared to those of eye lens crystallins alpha A and alpha
 CC B. The polypeptide shows extraordinarily high chaperone-like activity
 CC ranging between 3 - 6 times that of the eye lens crystallins alpha A and
 CC alpha B. The present sequence represents the chimera DNA that encodes
 CC human alpha BNAC crystallin
 XX
 SQ Sequence 531 BP; 90 A; 198 C; 130 G; 113 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.38e-80 Length: 531
 Score: 702.00 Matches: 137
 Percent Similarity: 85.96% Conservative: 15
 Best Local Similarity: 76.97% Mismatches: 19
 Query Match: 76.64% Indels: 6
 DB: 8 Gaps: 4

US-10-657-740-1 (1-173) x ABX12062 (1-531)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
 Db 1 ATGGACATCGCCATCCACCACCCCTGGATCGCGCCCTTCTTCTTCTTCCACTCCGCC 60
 Qy 20 SerArgLeuPheAspGlnPheGlyLeuPheGluTytrAspLeuPheProPhe 39
 Db 61 AGCCGCCCTTTGACAGTTCTTCGAGAGACCTTGGAGTCTGATCTTTTCCG--- 117
 Qy 40 LeuSerSerThrIleSerProTytrArgGln-----SerLeuPheArg-----Thr 55
 Db 118 ACGTCTACTTCCCTGAGTCCCTTCTTACCTTCGCCACCCCTCTTCTTCGCGGCCACCCAGC 177
 Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 Db 178 TGGTTTGACACTGGACTCTCAGAGATGCGCCTGGAGAGAGACAGGTTCCTCTCAACCTG 237
 Qy 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGlu 95
 Db 238 GATGTGAAGCACTTCTCCCGGAGACCTCACCGTAAGGTGCAGAGACACTTGTGGAG 297
 Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTytrIleSerArgGluPheHis 115
 Db 298 ATCCACGGAAGCACACAGAGCGCCAGAGCAGACACCGCTACATTTCCGTGAGTTCCAC 357
 Qy 116 ArgArgTytrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
 Db 358 CGCCGCTACCGCTCCGCTCCAACTGGACAGTCCGCGCTCTCTTGTCTCCCTGTCTGCC 417
 Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
 Db 418 GATGGCATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGGCCCTGGATGCCACCCGCC 477
 Qy 156 GluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
 Db 478 GAGCGAGCCATCCCGTGTGCGGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 531

RESULT 5
 ADQ78288

ID ADQ78288 standard; cDNA; 372 BP.
 XX
 AC ADQ78288;
 XX
 DT 04-NOV-2004 (first entry)
 XX
 DE Human N-terminal truncated alpha-crystallin DNA.
 XX
 KW alpha-crystallin; enhanced protein expression;
 KW enhanced protein secretion; protein aggregation; heat tolerance;
 KW elevated temperature; human; ds; gene.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 1..372
 FT /*tag= a
 FT /partial
 FT /product= "N-terminal truncated alpha-crystallin"
 FT /note= "No start codon given"
 XX
 PN US2004157289-A1.
 XX
 PD 12-AUG-2004.
 XX
 PF 08-SEP-2003; 2003US-00657740.
 XX
 PR 06-SEP-2002; 2002US-0408680P.
 XX
 PA (SALE/) SALERNO J C.
 PA (HANN/) HANNA M.
 PA (KORE/) KORETZ J F.
 PA (CRON/) CRONE D.
 PA (SMIT/) SMITH S M E.
 XX
 PI Salerno JC, Hanna M, Koretz JF, Crone D, Smith SME;
 XX
 DR WPI; 2004-580268/56.
 DR P-PSDB; ADQ78289.
 XX
 PT New truncated alpha-crystallin polypeptide derived from a wild-type alpha
 PT -crystallin protein, useful for enhancing protein (e.g. insulin or
 PT alcohol dehydrogenase) expression or secretion and for preventing protein
 PT aggregation.
 XX
 PS Claim 16; SEQ ID NO 2; 33pp; English.
 XX
 CC The invention relates to a truncated alpha-crystallin polypeptide derived
 CC from a wild-type alpha-crystallin protein, where the truncated
 CC polypeptide lacks an N-terminal sequence present in the wild-type
 CC protein. The composition and methods are useful for enhancing protein
 CC (e.g. insulin or alcohol dehydrogenase) expression or secretion and for
 CC preventing protein aggregation. These may also be used for creating a
 CC thermophilic host that tolerates elevated temperatures. The present
 CC sequence represents human N-terminal truncated alpha-crystallin DNA.
 XX
 SQ Sequence 372 BP; 66 A; 136 C; 98 G; 72 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.09e-72 Length: 372
 Score: 639.00 Matches: 123
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 69.76% Indels: 0
 DB: 13 Gaps: 0
 US-10-657-740-1 (1-173) x ADQ78288 (1-372)
 Qy 51 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLys 70
 Db 1 TCCCTCTTCCGACCGCTGCTGACTCCGCGCATCTCTGAGGTTCGATCCGACCGGACAAAG 60
 Qy 71 PheValIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGln 90

```

Db 61 TTGCTCATCTTCTCGATGTGAAGCATTCTCCCGGAGGACCTCACCGTGAAGTGCAG 120
Qy 91 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 110
Db 121 GACGACTTTGTGGAGATCCACGGAAGACACACGAGCGCCGACGACGACGCGTACATT 180
Qy 111 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 130
Db 181 TCCCGTAGTTCCACCGCGCTACCGCTGCGCTCCAGCTGAGCGACGTGGCCCTCTCT 240
Qy 131 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 150
Db 241 TGCTCCCTGCTGCGGATGGCATGCTGACCTTCTGTGCCCCCAAGATCCAGACTGCGCTG 300
Qy 151 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 170
Db 301 GATGCCACCCAGCGCGAGCGAGCCATCCCGTGTGCGGGAGGAGAGCCACCTCGGCT 360
Qy 171 ProSerSer 173
Db 361 CCTCTGCTCC 369

```

RESULT 6

ABZ35138
ID ABZ35138 standard; cDNA; 548 BP.

AC ABZ35138;

DT 05-FEB-2003 (first entry)

XX Human gene expression profile polynucleotide SEQ ID NO 250.

XX Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;
KW bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;
KW tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;
KW gene expression; gene; ss.

XX Homo sapiens.

XX WO200274979-A2.

XX 26-SEP-2002.

XX 20-MAR-2002; 2002WO-US008456.

XX 20-MAR-2001; 2001US-0276947P.

XX (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.

XX Wan J, Wang Y;

XX WPI; 2002-740862/80.

XX New gene expression profile generated from primary, endothelial,
PT epithelial, and muscle cell types, useful for identifying disease
PT pathologies involving alterations of gene expression, e.g. cancer.

XX Claim 10; Page 429; 850pp; English.

XX The invention relates to a gene expression profile comprising one or more
CC genes (ABZ34889-ABZ35692) and generated from a cell type. The cell type
CC is a coronary artery endothelium, umbilical artery or vein endothelium,
CC aortic endothelium, dermal microvascular endothelium, pulmonary artery
CC endothelium, myometrium microvascular endothelium, keratinocyte
CC epithelium, bronchial epithelium, mammary epithelium, prostate
CC epithelium, renal cortical epithelium, renal proximal tubule epithelium,
CC small airway epithelium, renal epithelium, umbilical artery smooth
CC muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,
CC dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,
CC aortic smooth muscle, mesangial cells, coronary artery smooth muscle,
CC bronchial smooth muscle, uterine smooth muscle, lung fibroblast,
CC osteoblasts or prostate stromal cell. The gene expression profile is used

CC for determining the level of RNA expression for a sample, determining the
CC phenotype of a cell and distinguishing cell types. The gene or a protein
CC expression profile is useful in identifying disease pathologies involving
CC alterations of gene expression. The assessment of expression profiles may
CC provide meaningful information with respect to tumour type and stage,
CC treatment methods, and prognosis. The gene or protein expression profile
CC may also be used for creating microarrays. The microarray is useful for
CC genetic and physical mapping of genomes, DNA sequencing, genetic or
CC medical diagnosis, genotyping of organisms, confirming cell or tissue
CC identifications and in identifying promising antibiotics, antiviral or
CC antifungal agents

XX Sequence 548 BP; 120 A; 178 C; 127 G; 123 T; 0 U; 0 Other;

Alignment Scores: 1.33e-53 Length: 548
Pred. No.: 493.00 Matches: 98
Score: 74.16% Conservative: 34
Best Local Similarity: 55.06% Mismatches: 34
Query Match: 53.82% Indels: 12
DB: 6 Gaps: 6

US-10-657-740-1 (1-173) x ABZ35138 (1-548)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 21 ATGGACATCGTATCCACCACCTTGGATCGCGCCCTTCTTCTTTTCACTCGGCC 80
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluLeuPheGluTyrAspLeuProPhe 39
Db 81 AGCGCGCTCTTTGACAGTCTCTCGGAGACACCTGTTGGAGTCTGATCTCTTCCA-- 137
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 138 ACTTCTACTTCTCCCTGAGCCCTTCTATCTTCGGCCACCTCTCATTCCTCGCGGACCCACG 197
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 198 TGGATTGACACTGGACTCTCAGATGCGCTGGAGAGGACAGGTTCTGTCAACCTG 257
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 258 GATGTGAAGCACTTCTCCCGAGAGGAGCTCAAGGTCAAAGTGTGGGTGATGTGATTGAG 317
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 318 GTGCACGGCAACATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGAGGATCCAC 377
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 378 AGGAATACCGGATCCCGAGCTGATGTGGACCTCTCACCATTACTTCACTCCCTGTCACT 437
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 438 GATGGGTCTCTCACTGTGAATGGACCAAGAACAA-----GCCCTCGGCCCA 485
Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 486 GAGCGCACCATTCCTCCATAACCCGTGAAGAGAGCGCTGTGTCTCACTGCGAGCCCCC 539

RESULT 7
ADZ46388
ID ADZ46388 standard; DNA; 528 BP.
XX
AC ADZ46388;
XX
DT 04-NOV-2004 (first entry)
XX Human lens crystalline protein coding sequence.
DE
XX Human lens crystalline protein coding sequence.
XX cytosolic; immunotoxin; cancer; mitochondrial malate dehydrogenase;
KW enzyme; human; MDH; ds; gene; lens crystalline protein.
XX

CC The invention relates to a hybrid protein chaperone for stabilising
 CC proteins and/or protein activities. Protein chaperones (also known as
 CC heat shock proteins) are divided into 4 families on the basis of their
 CC primary sequence and chaperone properties: HSP90, HSP70, HSP60 and sHSP
 CC (small heat shock protein). The invention is based upon the finding that
 CC among the sHSP family, which have a general structure of a central domain
 CC (called the alpha-crystallin domain) flanked by N and C-terminal regions,
 CC replacement of one or more regions of an sHSP with the corresponding
 CC region from a second sHSP can improve the activity compared to native
 CC sHSPs. In a particular embodiment of the invention, the hybrid chaperone
 CC is a hybrid sHSP designated alphaB-HSP27 comprising the N-terminus and
 CC central portion of alphaB-crystallin and the C-terminal tail of HSP27.
 CC However, the hybrid protein chaperones of the invention can comprise
 CC regions from HSP90, HSP70 and HSP60 families as well as from the sHSP
 CC family. The invention also encompasses methods for stabilising proteins
 CC such as enzymes, therapeutic proteins, diagnostic proteins, reporter
 CC proteins or antibodies, their fragments or conjugates in an aqueous
 CC solution using hybrid protein chaperones; stabilised protein formulation
 CC comprising at least one protein associated with the above hybrid protein
 CC chaperone; nucleic acids and vectors encoding a hybrid protein chaperone;
 CC and the coexpression of a recombinant protein of interest and a hybrid
 CC protein chaperone. The hybrid protein chaperones are useful as agents to
 CC prevent protein aggregation, as inhibitors of cell death and genome
 CC stability pathways, for identification of proteins that are in the
 CC process of unfolding, for the treatment of diseases involving altered
 CC protein conformations (e.g., cardiomyopathies, cataracts or
 CC neurodegenerative disease), or for the manufacture of a medicament for
 CC the treatment of such diseases. The present sequence represents a mutant
 CC human alphaB-crystallin DNA sequence in which the G at position 497 was
 CC mutated from T (see ADE75375) to generate a unique Aval restriction site
 CC which was used for subsequent cassette mutagenesis (along with a vector-
 CC based SacI site) to introduce heterologous C-terminal sequences onto
 CC alphaB-crystallin.

XX SQ Sequence 537 BP; 118 A; 171 C; 125 G; 123 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4, 22e-53 Length: 537
 Score: 489.00 Matches: 97
 Percent Similarity: 73.60% Conservative: 34
 Best Local Similarity: 54.49% Mismatches: 35
 Query Match: 53.38% Indels: 12
 DB: 10 Gaps: 6

US-10-657-740-1 (1-173) x ADE75374 (1-537)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
 DB 9 ATGGACATGCCATCCACACCCCTGGATCCGCGCCCTCTCTTCTTTCACATCCGCC 68
 QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
 DB 69 AGCGCGCTCTTTGACCACTTCTCGAGAGACACCTGTGGAGAGACCTGTGGAGTCTGTCCTCCCG--- 125
 QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
 DB 126 AGTCTACTTCTCTGAGTCCCTTCTACCTTCCGCCACCCCTCTTCTCGCGGCACCCAGC 185
 QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 DB 186 TGGTTTGACACTGGACTCTCAGAGATCGCCTGGAGAGACAGAGTCTCTCTGTCACCTG 245
 QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
 DB 246 GATGTGAAGCACTTCTCCCGCAGGAACTCAAAAGTTAAGGTGTGGAGATGTGATTGAG 305
 QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
 DB 306 GTGCTGGAACAAATCATGAAGCGCCAGATGATGAGATGATGATCTTCCAGGAGATTCCAC 365
 QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
 DB 366 AGGAATACCGGATCCACGCTGATGTAGACCCCTCTCACCATTACTTCTATCCCTGTCATCT 425

QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
 DB 426 GATGGGGTCTCCTCAGTGTGAATGACCAAGAAACAG-----GTCTCTGCCCT 473
 QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
 DB 474 GAGCGCACCATTTCCCATCCCGGAGAGAGAGCTGTGTCTACCGCACCCCC 527

RESULT 9

ADE75375

ID ADE75375 standard; DNA; 537 BP.

XX AC ADE75375;

DT 29-JAN-2004 (first entry)

XX Human wild-type alphaB-crystallin fragment-encoding DNA.

XX Hybrid protein chaperone; protein stabilisation; heat shock protein;
 KW sHSP family; protein aggregation inhibition; cell death inhibition;
 KW genome stability pathway inhibition; protein denaturation identification;
 KW protein conformation related disease; cardiomyopathy; cataract;
 KW neurodegenerative disease; cataract; ophthalmological; neuroprotective;
 KW gene therapy; alphaB-crystallin; wild-type; human; cassette mutagenesis;
 KW ds.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT mutation replace(497,G)
 FT /*tag= a

FT /note= "Mutagenesis of T to G at this site generates a
 unique Aval site between bases 493-498"

XX WO2003091266-A2.

XX 06-NOV-2003.

XX 23-APR-2003; 2003WO-GB001721.

XX 23-APR-2002; 2002GB-00009334.

XX (UYDU-) UNIV DUNDEE.

XX Quinlan R;

XX WPI; 2003-865571/80.

XX New hybrid protein chaperone (e.g. heat shock protein) useful for
 PT stabilising proteins and/or protein activities, or as an agent to prevent
 PT protein aggregation, or for treating diseases involving altered protein
 PT conformations.

XX Disclosure; Fig 12; 45pp; English.

XX The invention relates to a hybrid protein chaperone for stabilising
 CC proteins and/or protein activities. Protein chaperones (also known as
 CC heat shock proteins) are divided into 4 families on the basis of their
 CC primary sequence and chaperone properties: HSP90, HSP70, HSP60 and sHSP
 CC (small heat shock protein). The invention is based upon the finding that
 CC among the sHSP family, which have a general structure of a central domain
 CC (called the alpha-crystallin domain) flanked by N and C-terminal regions,
 CC replacement of one or more regions of an sHSP with the corresponding
 CC region from a second sHSP can improve the activity compared to native
 CC sHSPs. In a particular embodiment of the invention, the hybrid chaperone
 CC is a hybrid sHSP designated alphaB-HSP27 comprising the N-terminus and
 CC central portion of alphaB-crystallin and the C-terminal tail of HSP27.
 CC However, the hybrid protein chaperones of the invention can comprise
 CC regions from HSP90, HSP70 and HSP60 families as well as from the sHSP
 CC family. The invention also encompasses methods for stabilising proteins
 CC such as enzymes, therapeutic proteins, diagnostic proteins, reporter
 CC proteins or antibodies, their fragments or conjugates in an aqueous

CC solution using hybrid protein chaperones; stabilised protein formulation
 CC comprising at least one protein associated with the above hybrid protein
 CC chaperone; nucleic acids and vectors encoding a hybrid protein chaperone;
 CC and the coexpression of a recombinant protein of interest and a hybrid
 CC protein chaperone. The hybrid protein chaperones are useful as agents to
 CC prevent protein aggregation, as inhibitors of cell death and genome
 CC stability pathways, for identification of proteins that are in the
 CC process of unfolding, for the treatment of diseases involving altered
 CC protein conformations (e.g., cardiomyopathies, cataracts or
 CC neurodegenerative disease), or for the manufacture of a medicament for
 CC the treatment of such diseases. The present sequence represents a wild-
 CC type human alphaB-crystallin DNA sequence. The T at position 497 was
 CC mutated to G (see ADE75374) to generate a unique AwaI restriction site
 CC which was used for subsequent cassette mutagenesis (along with a vector-
 CC based SacI site) to introduce heterologous C-terminal sequences onto
 CC alphaB-crystallin.

SQ Sequence 537 BP; 118 A; 171 C; 124 G; 124 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.22e-53 Length: 537
 Score: 489.00 Matches: 97
 Percent Similarity: 73.60% Conservativity: 34
 Best Local Similarity: 54.49% Mismatches: 35
 Query Match: 53.38% Indels: 12
 DB: 10 Gaps: 6

US-10-657-740-1 (1-173) x ADE75375 (1-537)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
 Db 9 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTCTTCTTCTTCCATCCCTCCC 68
 Qy 20 SerArgLeuPheAspGlnPheGlyLeuPheGluGlyLeuPheGluTyrAspLeuPhePhe 39
 Db 69 AGCGCCCTCTTGGACATCTTCGGAGAGACCTGTGGAGTGTGATCTTTCCCG--- 125
 Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
 Db 126 ACGTCTACTCTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTCTCTCGCGCCACCCAGC 185
 Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 Db 186 TGGTTGACATGGACTCTGAGATGCGCTGGAGAGACAGAGTCTCTGTCAACCTG 245
 Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
 Db 246 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAGGTGGGAGATGTGATGAG 305
 Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
 Db 306 GTGCATGGAAACATGAAGAGCGCCAGCATGAACATGTTTCATCTCCAGGGAGTTCCAC 365
 Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
 Db 366 AGGAATACCGGATCCCGATGATGATAGACCTCTCACCATTTACTTCACTCTGTCACT 425
 Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
 Db 426 GATGGGTCTTCACTGTGAATGGACCAAGAAACAG-----GTCTTGGCCCT 473
 Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
 Db 474 GAGCGCACCATTTCCCATCACCCTGAGAGAGAGAGCTGTGTGTCACCGCAGCCCCC 527
 RESULT 10
 AAX39668
 ID AAX39668 standard; DNA; 691 BP.
 XX AC AAX39668;
 XX AC AAX39668;
 XX DT 02-JUL-1999 (first entry)
 XX

DE Renal cancer associated gene.

XX Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer; ss.

OS Homo sapiens.

PN WO9904265-A2.

XX 28-JAN-1999.

XX 15-JUL-1998; 98WO-US014679.

PR 17-JUL-1997; 97US-00896164.

PR 10-OCT-1997; 97US-0061599P.

PR 10-OCT-1997; 97US-0061765P.

PR 10-OCT-1997; 97US-00948705.

PR 11-OCT-1997; 97GB-00021697.

PR 22-JUN-1998; 98US-00102322.

XX (LUDW-) LUDWIG INST CANCER RES.

PA Old LJ, Scanlan MJ, Stockert B, Gure A, Chen Y, Gout I;
 PI O'hare M, Obata Y, Pfreundschuh M, Tureci O, Sahin U;
 DR WPI; 1999-132448/11.

XX New isolated cancer associated nucleic acids and polypeptides - isolated
 PT using sera from cancer patients, used to develop products for the
 PT diagnosis, monitoring or treatment of cancers.

PS Claim 67; Page 478; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer

SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.15e-53 Length: 691
 Score: 489.00 Matches: 97
 Percent Similarity: 73.60% Conservativity: 34
 Best Local Similarity: 54.49% Mismatches: 35
 Query Match: 53.38% Indels: 12
 DB: 2 Gaps: 6

US-10-657-740-1 (1-173) x AAX39668 (1-691)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19

Db 26 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTCTTCTTCTTCCATCCCTCCC 85

Qy 20 SerArgLeuPheAspGlnPhePheGlyLeuPheGluTyrAspLeuPhePhe 39

Db 86 AGCGCCCTCTTGGACATCTTCGGAGAGACACTGTGGAGTGTGATCTTTCCCG--- 142

Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55

Db 143 ACGTCTACTCTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTTCTTCCGCGGACCCAGC 202


```

Db 203 TGGTTGACACTGGACTCTCAGAGATGCGCTGGAGAAGGACAGGTTCTCTGTCAACCTG 262
Qy 76 AspVallyHisPheSerProGluAspLeuThrVallyValGlnAspPheValGlu 95
Db 263 GATGGAAGCACTCTCCCGAGAGAACTCAAGTTAAGTGTGGGAGATGTGATTGAG 322
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGAAACATGAAGAGCGCCAGCATGAACATGGTTTTCATCTCCAGGAGTTCCAC 382
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 383 AGGAATACCGATCCGACTGATGAGACCTCTCACCATTACTTCACTCTGTCATCT 442
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGTCTCTCACTGTGTAATGGACCAAGAAACAG-----GTCTCTGGCCCT 490
Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTCCTCCATCACCCTGAGAGAGAGAGCTGTGTGCACGCGAGCCGCC 544

```

RESULT 12

```

ABL62427
ID ABL62427 standard; DNA; 691 BP.
XX
AC ABL62427;
XX
DT 15-MAY-2002, (first entry)
XX
XX Colon adenocarcinoma related gene sequence SEQ ID NO:764.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytosarctic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX

```

```

PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233133P.
PR 20-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 22-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0234927P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236103P.
PR 28-SEP-2000; 2000US-0236111P.

```

```

PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 02-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, and determining a change
XX in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 764; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an anti-
XX neoplastic agent. The method involves exposing cells to a chemical agent
XX to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening an
XX anti-neoplastic agent, and can be used for producing a product which is
XX the data collected with respect to the anti-neoplastic agent as a result
XX of M1, and the data is sufficient to convey the chemical structure and/or
XX properties of the agent. M1 can be used in the treatment of cancer such
XX as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
XX prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
XX cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
XX cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
XX tumour
XX
XX SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred. No.: 6.15e-53 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 6 Gaps: 6

```

```

US-10-657-740-1 (1-173) x ABL62427 (1-691)
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGGACATCGCCATCCACCACCCCTGGATCGCGCCCTCTTTCTTTCTTCCACTCCGCC 85
Qy 20 SerArgLeuPheAspGlnPhePheGlyGlyGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCTCTTTGACAGTCTTCGGAGAGACCTGTGGAGTCTGATCTTTTCCG--- 142
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 143 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCTCTCTCTCTCGGGCACCAGC 202
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75

```



```
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGACATCGCATCCACACCCCTGGATCGCGCCCTCTTCTTCTTCCACTCCCCC 85
Qy 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCTCTTTGACCACTCTTCGAGAGCACCTGTTGGAGTCTGATCTTTTCCCG--- 142
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 143 ACGTCTACTTCCCTGAGTCCCTTACCTTCGGCCACCCCTCTTCTTCTGCGGCGACCCAGC 202
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 203 TGGTTTGACACTGCACTCTCAGATGCGCTCGGAGAGGACAGTCTCTCTGCAACCTG 262
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGGAACAACATGAAGAGCGCCAGATGAACATGCTTTCATCTCCAGGAGTTCCAC 382
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 393 AGGAATACCGATCCAGCTGATGATAGACCCCTCACCATTACTTCACTCTGTCATCT 442
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGTCTCTCCTGTAATGGACCAAGAAACAG-----GTCTCTGGCCCT 490
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTTCCCATCACCCTGGAAGAGAGCCCTGCTGTCCCGAGCCCCC 544
RESULT 16
ID ADN05800 standard; cDNA; 691 BP.
AC ADN05800;
DT 01-JUL-2004 (first entry)
XX Antipsoriatic cDNA sequence #1131.
DE Antipsoriatic cDNA sequence #1131.
KW ds; gene; antipsoriatic; gene therapy; psoriasis; diagnosis.
XX Homo sapiens.
XX WO2004028479-A2.
XX 08-APR-2004.
XX 25-SEP-2003; 2003WO-US030907.
XX 25-SEP-2002; 2002US-0414006P.
XX (GETH ) GENENTECH INC.
XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
PI Wu TD;
XX WPI; 2004-305105/28.
XX P-PSDB; ADN05801.
XX New PRO nucleic acid or polypeptide, useful for preparing a
PT pharmaceutical composition for diagnosing or treating psoriasis in a
PT mammal.
XX Claim 1; SEQ ID NO 2194; 3069pp; English.
XX
```

```
CC The invention relates to novel polynucleotide and polypeptides for
CC treating psoriasis or a sequence having at least 80% identity to the
CC above sequences. The nucleic acid is useful for preparing a composition
CC for diagnosing or treating psoriasis in a mammal. This sequence
CC corresponds to one of the polynucleotides of the invention.
XX
SQ Sequence 691 BP; 168 A; 208 C; 148 G; 167 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.15e-53 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 12 Gaps: 6

US-10-657-740-1 (1-173) x ADN05800 (1-691)
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGACATCGCATCCACACCCCTGGATCGCGCCCTCTTCTTCTTCCACTCCCCC 85
Qy 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCTCTTTGACCACTCTTCGAGAGCACCTGTTGGAGTCTGATCTTTTCCCG--- 142
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 143 ACGTCTACTTCCCTGAGTCCCTTCTTCTTCGGCCACCCCTCTTCTTCTGCGGCGACCCAGC 202
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 203 TGGTTTGACACTGCACTCTCAGATGCGCTCGGAGAGGACAGTCTCTCTGCAACCTG 262
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 263 GATGGAAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAG 322
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGGAACAACATGAAGAGCGCCAGATGAACATGCTTTCATCTCCAGGAGTTCCAC 382
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 393 AGGAATACCGATCCAGCTGATGATAGACCCCTCACCATTACTTCACTCTGTCATCT 442
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGTCTCTCCTGTAATGGACCAAGAAACAG-----GTCTCTGGCCCT 490
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTTCCCATCACCCTGGAAGAGAGCCCTGCTGTCCCGAGCCCCC 544
RESULT 17
ID ADR24825 standard; DNA; 691 BP.
XX ADR24825;
AC ADR24825;
DT 21-OCT-2004 (first entry)
XX Breast cancer prognosis marker #686.
XX ds; breast cancer; prognosis; gene expression; diagnosis.
XX Homo sapiens.
XX WO2004065545-A2.
XX 05-AUG-2004.
XX 15-JAN-2004; 2004WO-US001100.
XX
```


XX 26-FEB-1999; 99US-0122487P.
XX (GSET) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 14936; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 911 BP; 221 A; 260 C; 211 G; 219 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 9,3e-53 Length: 911
XX Score: 489.00 Matches: 97
XX Percent Similarity: 73.60% Conservatives: 34
XX Best Local Similarity: 54.4% Mismatches: 35
XX Query Match: 53.38% Indels: 12
XX DB: 3 Gaps: 6
XX
XX US-10-657-740-1 (1-173) x AAC10863 (1-911)
XX
XX 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
XX 284 ATGGACATCGCCATCCACCACCCCTGGATCGCGCCCTTCTTCTTCCACTCCGCC 343
XX
XX 20 SerArgLeuPheAspGlnPheGlyLeuPheGluTyrAspLeuLeuProPhe 39
XX 344 AGCCGCCCTTTGACACAGTTCTTCGGAGAGACCTGTGGAGTCTGTCTTCCCG--- 400
XX
XX 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
XX 401 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCCACCCCTCTCTCTCGGGCACCACG 460
XX
XX 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
XX 461 TGGTTTGACACTGGAGTCTCAGAGATGCGCTCGAGAGACAGGTCTCTGTCAACCTG 520
XX
XX 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
XX 521 GATGTGAAGCACTTCTCCCGAGAGAACTCAAAGTTAAGGTGTTGGGAGATGTGATTGAG 580
XX
XX 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
XX 581 GTGCATGGAAACATGAAGAGCCCGCAGATGATGATGTTTCATCTCCAGGGAGTTCCAC 640
XX
XX 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
XX 641 AGGAAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTACTTCTATCTGTCATCT 700
XX
XX 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
XX 701 GATGGGGTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCT 748

Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 749 GAGCGCACCATTCCTCCATCACCCTGTGAGAGAAGCTGTGTCTCAGCGCAGCCCC 802
RESULT 21
AAC10861
ID AAC10861 standard; cDNA; 927 BP.
XX AC AAC10861;
XX 06-OCT-2000 (first entry)
XX DE. Human secreted protein 5' EST, SEQ ID NO: 14936.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GSET) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 14936; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 927 BP; 230 A; 267 C; 224 G; 206 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 9,54e-53 Length: 927
XX Score: 489.00 Matches: 97
XX Percent Similarity: 73.60% Conservatives: 34
XX Best Local Similarity: 54.49% Mismatches: 35
XX Query Match: 53.38% Indels: 12
XX DB: 3 Gaps: 6
XX
XX US-10-657-740-1 (1-173) x AAC10861 (1-927)
XX
XX 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
XX 300 ATGGACATCGCCATCCACCACCCCTGGATCGCGCCCTTCTTCTTCCACTCCGCC 359
XX
XX 20 SerArgLeuPheAspGlnPheGlyLeuPheGluTyrAspLeuLeuProPhe 39
XX 701 GATGGGGTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCT 748

Db 360 AGCCGCTCTTTGACCACTTCTTCGGAGACACCTGTTGGAGTCTGTATCTTTTCCCG--- 416
Qy 40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 417 ACCTCTACTCTCCCTGAGCCCTTCTACCTTCGGCCACCCTCTCTCTCGGGCACCACG 476
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 477 TGGTTTGACATGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 536
Qy 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGlu 95
Db 537 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAAGTTGTTGGGAGATGTGATTGAG 596
Qy 96 IleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 597 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGTTTTCATCTCCAGGAGGTTCCAC 656
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 657 AGGAATACCGGATCCGAGCTGATGATAGACCTCTCACCATTACTTCTATCTCTCATCT 716
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 717 GATGGGTCTCTCACTGTGAATGACCAAGGAAACAG-----GTCTCTGGCCCT 764
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 765 GAGCGCACCATTTCCCATCCCGTGAAGAGAGCGCTGTCTCAGCGCAGCCCC 818
RESULT 22
AAC03892
ID AAC03892 standard; cDNA; 942 BP.
XX AC AAC03892;
XX
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 3890.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX (G8ST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
DR P-PSDB; AAG03886.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 3890; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
XX vectors
XX
SQ Sequence 942 BP; 223 A; 267 C; 224 G; 228 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 9.77e-53 Length: 942
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 3 Gaps: 6
US-10-657-740-1 (1-173) x AAC03892 (1-942)
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 315 ATGGACATCGCCATCCACACCCCTGGATCGCGCCCTCTTCTTCTTCCATCCCTCCC 374
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGlyTyrAspLeuProphe 39
Db 375 AGCGGCTCTTTGACCACTTCTTCGGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 431
Qy 40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 432 ACGTCTACTCTCCCTGAGTCCCTTCTACCTTCGGCCACCCTCTCTCTCGGGCACCACG 491
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 492 TGGTTTGACATGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 551
Qy 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGlu 95
Db 552 GATGTGAAGCACTTCTCCCGAGAGAACTCAAGTTAAAGTTGTTGGGAGATGTGATTGAG 611
Qy 96 IleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 612 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGTTTTCATCTCCAGGAGGTTCCAC 671
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 672 AGGAATACCGGATCCCGAGCTGATGATAGACCTCTCACCATTACTTCTCTCTGTCTCT 731
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 732 GATGGGTCTCTCACTGTGAATGAGCAAGGAAACAG-----GTCTCTGGCCCT 779
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 780 GAGCGCACCATTTCCCATCCCGTGAAGAGAGCGCTGTCTCAGCGCAGCCCC 833
RESULT 23
AAZ33574/C
ID AAZ33574 standard; cDNA; 1036 BP.
XX AC AAZ33574;
XX
XX 08-DEC-1999 (first entry)
XX Human breast tumour-associated EST 34.
XX
XX Expressed sequence tag; EST; human; breast; cancer; cytostatic;
KW medicaments; gene therapy; treatment; fat metabolism; ss.
XX Homo sapiens.
XX OS
XX DE19813835-A1.
XX FN
XX


```
Alignment Scores:
Pred. No.: 1,358-52 Length: 528
Score: 485.00 Matches: 96
Percent Similarity: 74.16% Conservative: 36
Best Local Similarity: 53.93% Mismatches: 34
Query Match: 52.95% Indels: 12
DB: 10 Gaps: 6

US-10-657-740-1 (1-173) x AD852642 (1-528)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 1 ATGGACATAGCATCCACCACCCCTGGATCCGGCGTCCCTCTTCTTCCACTCCCA 60
QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 61 AGCCGCCCTCTTTGACCACTTCTTCGAGAGACACCTGTTGGAGCTGACCTCTTCTCT--- 117
QY 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55
DB 118 ACAGCCACTTCCCTGAGCCCTTCTACCTTCGGCCACCCCTCTTCTTGGGGCACCCTAGC 177
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 178 TGGATTGACATGGGCTCTCAGAGATCGTATGGAGAGACAGGTTCTCTGTGAACCTG 237
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 238 GACGTGAAGCACTTCTCCAGAGAACTCAAGTCAAGTCTTCGGAGACGTGATTGAG 297
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 298 GTGCACGGCAAGCAGAGCGCAGGACGAACTGGCTTCATCTCAGAGGAGTTCCAC 357
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 358 AGGAAGTACCGGATCCAGCCGACGCTGGATCCTCTCACCATTAATCTTCTCCCTGTCATCG 417
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 418 GATGAGTCTCTCACTGTGAATGGACCAAGAAACAG-----GCCCTCGGCCCT 465
QY 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
DB 466 GAGCGCACCATTCATCCATCCCGTGAAGAGAACGCTGTGTCTGCTGACGACCCCT 519

RESULT 25
ID ABT41872
XX ABT41872 standard; DNA; 528 BP.
AC ABT41872;
XX
XX
DT 26-JUN-2003 (first entry)
XX
XX
DE Toxicity modelling related rat gene SEQ ID No 1574.
XX
XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW database; drug screening; toxicity assay; rat; ds.
XX
XX Rattus norvegicus.
XX
XX WO200295000-A2.
XX
XX 28-NOV-2002.
XX
XX 22-MAY-2002; 2002WO-US016173.
XX
XX 22-MAY-2001; 2001US-0292335P.
XX 13-JUN-2001; 2001US-0297523P.
XX 19-JUN-2001; 2001US-0298925P.
XX 10-JUL-2001; 2001US-0303807P.
XX 10-JUL-2001; 2001US-0303808P.
```

```
PR 10-JUL-2001; 2001US-0303810P.
PR 28-AUG-2001; 2001US-0315047P.
PR 27-SEP-2001; 2001US-0324928P.
PR 22-OCT-2001; 2001US-0330462P.
PR 01-NOV-2001; 2001US-0330867P.
PR 21-NOV-2001; 2001US-0331805P.
PR 06-DEC-2001; 2001US-0336144P.
PR 19-DEC-2001; 2001US-0340873P.
PR 21-FEB-2002; 2002US-0357842P.
PR 21-FEB-2002; 2002US-0357843P.
PR 21-FEB-2002; 2002US-0357844P.
PR 15-MAR-2002; 2002US-0364134P.
PR 08-APR-2002; 2002US-0370144P.
PR 08-APR-2002; 2002US-0370206P.
PR 08-APR-2002; 2002US-0370247P.
PR 17-APR-2002; 2002US-0372794P.
PR 21-APR-2002; 2002US-0371679P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
PI WPI; 2003-148464/14.
XX
XX Predicting at least one toxic effect of a compound, useful for toxicity
XX modeling, comprises preparing a gene expression profile of a tissue or
XX cell sample exposed to the compound, and comparing the gene expression
XX profile to a database.
XX
XX Example 4; Page; 446pp; English.
XX
XX The invention relates to a novel method of predicting at least one toxic
XX effect of a compound. The method comprises a gene expression profile of a
XX tissue or cell sample exposed to the compound, and comparing the gene
XX expression profile to a database comprising at least part of the data or
XX information given in the specification. The methods are useful for
XX predicting at least one toxic effect of a compound, predicting the
XX progression of a toxic effect of a compound, predicting the renal
XX toxicity of a compound, or identifying toxicity markers in tissues or
XX cells exposed to known renal toxin. The genes are useful as toxicity
XX markers in drug screening and toxicity assays, in monitoring disease or
XX physiological states, or disease progression. This polynucleotide
XX represents a rat DNA sequence relating to the toxic effect database
XX described in a rat DNA specification. NOTE: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from the World Intellectual Property
XX Organization
XX
XX SQ Sequence 528 BP; 114 A; 170 C; 128 G; 116 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 1,358-52 Length: 528
Score: 485.00 Matches: 96
Percent Similarity: 74.16% Conservative: 36
Best Local Similarity: 53.93% Mismatches: 34
Query Match: 52.95% Indels: 12
DB: 10 Gaps: 6
```

```
US-10-657-740-1 (1-173) x ABT41872 (1-528)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 1 ATGGACATAGCATCCACCACCCCTGGATCCGGCGTCCCTCTTCTTCCACTCCCA 60
QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 61 AGCCGCCCTCTTTGACCACTTCTTCGAGAGACACCTGTTGGAGCTGACCTCTTCTCT--- 117
QY 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55
DB 118 ACAGCCACTTCCCTGAGCCCTTCTACCTTCGGCCACCCCTCTTCTTGGGGCACCCTAGC 177
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
```

```

Db 178 TGGATTGACATGGGCTCTCAGAGATCGTATGGAGAGGACAGGTTCTCTGTGAACCTG 237
Qy 76 AspVallyHisPheSerProGluAspLeuThrVallyValGlnAspPheValGlu 95
Db 238 GACGTGAAGCACTTCTCCAGAGAACTCAAGTCAAGGTTCTGGGAGAGCTGATTGAG 297
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 298 GTGACGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 357
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 358 AGGAAGTACCGGATCCAGCCGACGATGATCCTCTCACCATTACTTCTCTCTGTGTCATCG 417
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 418 GATGAGTCTCTCACTGTGAATGGACCAAGAAACAG-----GCTCTGGCCCT 465
Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 466 GAGGCGACCATTCCTCATCCCGTGAAGAGAGGCTGTGTCACTGTCAGGCCCT 519

RESULT 26
ABK63512
ID ABK63512 standard; cDNA; 1247 BP.
XX
AC ABK63512;
XX
DT 18-JUN-2002 (first entry)
XX
DE Rat sequence differentially expressed in response to a hepatotoxin #1419.
XX
KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
XX differential expression; centrilobular necrosis; steatosis.
XX
OS Rattus norvegicus.
XX
FN WO200210453-A2.
XX
PD 07-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-US023872.
XX
PR 02-NOV-2000; 2000US-0222040P.
XX
PR 11-MAY-2001; 2001US-0290029P.
XX
PR 15-MAY-2001; 2001US-0290645P.
XX
PR 22-MAY-2001; 2001US-0292336P.
XX
PR 06-JUN-2001; 2001US-0295798P.
XX
PR 13-JUN-2001; 2001US-0297457P.
XX
PR 19-JUN-2001; 2001US-0298884P.
XX
PR 09-JUL-2001; 2001US-0303459P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter MW, Johnson KR, Castile AL, Elashoff MR;
XX
DR WPI; 2002-241625/29.
XX
PT Predicting toxic effects of compounds or the progression of these toxic
XX effects by determining the changes in gene expression in tissues or cells
XX exposed to the toxin and comparing these to gene expression in unexposed
XX tissues or cells.
XX
PS Claim 1; SEQ ID NO 1419; 239pp; English.
XX
CC The invention relates to methods for predicting toxic effects of
XX compounds or the progression of these toxic effects by determining the
XX global changes in gene expression in tissues or cells exposed to the
XX toxin and comparing these to gene expression in unexposed tissues or
XX cells. Also included are methods of predicting at least one toxic effect
XX of a compound or progression of a toxic effect, preferably the

```

```

CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of
CC the genes is indicative of at least one toxic effect or progression. The
CC method can also be used to identify an agent which modulates the toxic
CC response and predict cellular pathways that a compound modulates in a
CC cell. The methods utilise a set of at least two probes (on a solid
CC support in kit form), where each of the probes comprises a sequence that
CC specifically hybridises to a gene listed in the specification, a computer
CC system comprising a database containing information identifying the
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC set of genes comprising at least two genes listed in the specification,
CC and a user interface to view the information used to present information
CC identifying the expression level in a tissue or cell of at least one gene
CC listed in the specification. The method is useful for elucidating global
CC changes in gene expression and for identifying toxicity markers in
CC tissues or cell exposed to a known toxin. The genes may be used as
CC toxicity markers in drug screening and toxicity assays. The genes and
CC gene expression information may be used as diagnostic markers for the
CC prediction or identification of the physiological state of tissue or cell
CC sample that has been exposed to a compound or agent. Hepatotoxicity is
CC characterised by centrilobular necrosis and steatosis. The present
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC which is differentially expressed in response to a hepatotoxic agent
XX
SQ Sequence 1247 BP; 294 A; 360 C; 306 G; 287 T; 0 U; 0 Other;

```

```

Alignment Scores:
Pred No.: 4.88e-52 Length: 1247
Score: 485.00 Matches: 96
Percent Similarity: 74.16% Conservative: 36
Best Local Similarity: 53.93% Mismatches: 34
Query Match: 52.95% Indels: 12
DB: 6 Gaps: 6

```

US-10-657-740-1 (1-173) x ABK63512 (1-1247)

```

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 583 ATGGACATAGCATCCACCCCTGGATCGGGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 642
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPheProPhe 39
Db 643 AGCGCCTCTTTGACCACTTCTTCGGAGAGACCTGTTGGAGTCTGACCTCTCTCTCTCTCT 699
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 700 ACAGCCACTTCCCTGAGCCCTCTTACTCTGCGCACCTCTCTCTCTCTCTCTCTCTCTCTCT 759
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 760 TGGATTGACACTGGGCTCTCAGAGATCGTATGGAAGAGACAGGTTCTCTGTGAACCTG 819
Qy 76 AspVallyHisPheSerProGluAspLeuThrVallyValGlnAspAspPheValGlu 95
Db 820 GACGTGAAGCACTTCTCTCCAGAGAACTCAAGTCTCTGGGAGAGCTGATTGAG 879
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 880 GTGACGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 939
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 940 AGGAAGTACCGGATCCAGCCGACGATGATCCTCTCACCATTACTTCTCTCTCTCTCTCTCT 999
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 1000 GATGAGTCTCTCACTGTGAATGGACCAAGAAACAG-----GCTCTGGCCCT 1047
Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
Db 1048 GAGGCGACCATTCCTCATCCCGTGAAGAGAGGCTGTGTCACTGTCAGGCCCT 1101

```


XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors
 XX SQ Sequence 913 BP; 228 A; 257 C; 220 G; 202 T; 0 U; 6 Other;

Alignment Scores:
 Pred. No.: 1.08e-48 Length: 913
 Score: 457.50 Matches: 94
 Percent Similarity: 71.91% Conservativity: 34
 Best Local Similarity: 52.81% Mismatches: 33
 Query Match: 49.95% Indels: 17
 DB: 3 Gaps: 7

US-10-657-740-1 (1-173) x AAC10862 (1-913)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
 Db ATGGACATCGGCATCCACCCCTTC-----TTKCCCTTTCATCTCCCCC 345
 Qy 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
 Db AGCCGCCTCTTGGACCACTCTTCGCGAGCACCTGTTGGAGTCKGATSTTTCCG--- 402
 Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
 Db ACGTCTACTCTCCCTGAGTGCCTTCTACCTTCGGCCACCCTCTCTCGGGGACCCACG 462
 Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 Db TGGTTTGACATGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 522
 Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
 Db GATGTGAAGCACTTCTCCCGAGGAACTCAAGTTAGGTGTTGGGAGATGTGATTGAG 582
 Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
 Db GTGCATGGAAACATGAAGAGCGCCAGCATGATGATGATGATGATGATGATGATGATG 642
 Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
 Db AGGAATACCGGATCCCGATGATGATGATGATGATGATGATGATGATGATGATGATG 702
 Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
 Db GATGGGTCTCTCACTGTGTAATGGACCAAGAAACAG-----GTCTCTGGCCCT 750
 Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 171
 Db GAGCGCACCATTCCTCCATCACCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 804

RESULT 29

ABX39112

ID ABX39112 standard; cDNA; 449 BP.

XX ABX39112;

AC ABX39112;

DT 20-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #4277.

XX Bovine; SB; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX Bos Taurus.
 OS US2002137139-A1.
 XX 26-SEP-2002.
 XX 24-SEP-2001; 2001US-00960352.
 XX 12-JAN-1999; 99US-0115707P.
 PR 11-JAN-2000; 2000US-00480902.
 PA (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TXON/) TAO N.
 PA (WARR/) WARREN W C.
 BYatt JC, Mathialagan N, Tao N, Warren WC;
 WPI; 2003-110599/10.
 PT New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and analysis,
 PT cattle breeding, or for genetically improving cattle.
 PS Claim 2; SEQ ID NO 4277; 245pp; English.
 XX The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
 CC appearing as ABX34836-BX49947, or complements of them. Also included are
 CC (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non-translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?docID=20020137139
 XX SQ. Sequence 449 BP; 91 A; 148 C; 100 G; 110 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.66e-44 Length: 449
 Score: 418.00 Matches: 79
 Percent Similarity: 77.30% Conservativity: 30
 Best Local Similarity: 56.03% Mismatches: 26
 Query Match: 45.63% Indels: 6
 DB: 8 Gaps: 4

US-10-657-740-1 (1-173) x ABX39112 (1-449)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
 Db ATGGATATCGGCATCTCCACCCCTCGATCGCGCCCTTCTTCCTCCCTTCCACTCTCCC 89

QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
 DB AGCCGCTTTCAGCAGTTTTCGGGAGGACCTGTAGAGTCTGATCTCTCCCA--- 146
 QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
 DB GCTTCTACTCTCCGAGCCCTTCTACCTTCGGCGCCCTCATTTCTGGGGCACCACG 206
 QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 DB TGGATTGACACTGGGCTCTCAGAGATGCTCTGGAGAAGGACAGATTCTCTGTCAACCTG 266
 QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
 DB GATGTGAAGCACTTCTCCCGAGGAACCTCAAGGTCAAGGTCTGGGAGATGTGATTGAG 326
 QY 96 IleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115
 DB GTGCATGGCAACATGAAGAGCGCCAGGATGAACATGGTTTATCTCCGGGAGTTCCAC 386
 QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
 DB AGGAATAACCGGATCCGAGTACGCTGACGTGAACCTCTCGCCATTACTCTCATCTGTCGTCT 446
 QY 136 Asp 136
 DB 447 GAT 449
 RESULT 30
 ABX47457
 ID ABX47457 standard; cDNA; 396 BP.
 AC ABX47457;
 XX
 XX 21-FEB-2003 (first entry)
 XX
 XX Bovine EST associated with lactation/muscle/fat deposition #12622.
 XX
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 XX US2002137139-A1.
 XX
 PD 26-SEP-2002.
 XX
 XX 24-SEP-2001; 2001US-00960352.
 XX
 PR 12-JAN-1999; 99US-0115707P.
 PR 11-JAN-2000; 2000US-00480902.
 XX
 PA (BYATT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;
 XX WPI; 2003-110599/10.
 XX
 PT New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and analysis,
 PT cattle breeding, or for genetically improving cattle.
 XX
 PS Claim 2; SEQ ID NO 12622; 245pp; English.
 XX
 XX The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,

CC appearing as ABX34836-ABX49947, or complements of them. Also included are
 CC : (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non-translated sequence that
 CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?docID=20020137139
 XX
 SQ Sequence 396 BP; 80 A; 127 C; 93 G; 96 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.82e-41 Length: 396
 Score: 394.00 Matches: 74
 Percent Similarity: 77.52% Conservative: 26
 Best Local Similarity: 57.36% Mismatches: 23
 Query Match: 43.01% Indels: 6
 DB: 8 Gaps: 4

US-10-657-740-1 (1-173) x ABX47457 (1-396)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
 DB 13 ATGGATATCGCATCCACCACCCCTGGATCGCGCCCTCTTCCCTTTCCTCTCCCTCTCC 72
 QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
 DB 73 AGCGGCTCTTTGACCAAGTTTTCGGGAGCACCCTGTTGGAGTCTGATCTCTTCCCA--- 129
 QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
 DB 130 GCTTCTACTCTCCCTGAGCCCTTCTACCTTCGGCGCCCTCATTTCTGGGGCACCACG 189
 QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 DB 190 TGGATTGACACTGGGCTCTCAGAGATGCTCTGGAGAAGGACAGATTCTCTGTCAACCTG 249
 QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
 DB 250 GATGTGAAGCACTTCTCCCGAGGAACCTCAAGGTCAAGGTCTGGGAGATGTGATTGAG 309
 QY 96 IleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115
 DB 310 GTGCATGGCAACATGAAGAGCGCCAGGATGAACATGGTTTATCTCTCCGGGAGTTCCAC 369
 QY 116 ArgArgTyrArgLeuProSerAsnVal 124
 DB 370 AGGAATAACCGGATCCCGAGCTGACGTG 396

RESULT 31
 ACH93386
 ID ACH93386 standard; DNA; 211 BP.
 XX
 XX ACH93386;
 XX
 XX 29-JUL-2004 (first entry)
 XX
 XX Human genome derived single exon probe #26581.
 DE
 XX

KW Human; probe: ss; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX
 OS Homo sapiens.
 XX
 PN US2003194704-A1.
 XX
 XX 16-OCT-2003.
 XX
 PF 03-APR-2002; 2002US-00029386.
 XX
 XX 03-APR-2002; 2002US-00029386.
 PR
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK;
 XX
 XX WPI; 2004-119264/12.
 DR
 XX New human genome-derived single exon nucleic acid probes useful for human
 XX gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 PS Claim 1; SEQ ID NO 26581; 80pp; English.
 XX
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridises under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subsequence, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above). The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterising gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docID=20030194704
 XX
 SQ Sequence 211 BP; 32 A; 85 C; 55 G; 39 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,73e-38	Length:	211
Score:	367.00	Matches:	70
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.07%	Indels:	0
DB:	12	Gaps:	0

US-10-657-740-1 (1-173) x ACH93386 (1-211)
 QY 104 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAan 123
 |||||
 Db 1 CAGGACGACACACGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCCGTCACAC 60
 |||||
 QY 124 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 143
 |||||
 Db 61 GTGGACCAAGTCGGCCCTCTCTTGCTCCCTGTCGCGATGCGATGCTGACCTTCTGTGGC 120
 |||||
 QY 144 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 163
 |||||
 Db 121 CCCAAGATCCAGACTGGCTGAGTGCACCCACCGCGAGCGAGCCATCCCGGTGTGGCG 180
 |||||
 QY 164 GluGlyLysProThrSerAlaProSerSer 173
 |||||
 Db 181 GAGGAGAGCCACCTCGGCTCCCTCGTCC 210
 |||||

RESULT 32

ACH79686
 ID ACH79686 standard; DNA; 573 BP.
 XX
 AC ACH79686;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Human genome derived single exon probe #12881.
 XX
 KW Human; probe: ss; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX
 OS Homo sapiens.
 XX
 PN US2003194704-A1.
 XX
 PD 16-OCT-2003.
 XX
 PF 03-APR-2002; 2002US-00029386.
 XX
 XX 03-APR-2002; 2002US-00029386.
 PR
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK;
 XX
 XX WPI; 2004-119264/12.
 DR
 XX New human genome-derived single exon nucleic acid probes useful for human
 XX gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 PS Claim 15; SEQ ID NO 12881; 80pp; English.
 XX
 CC The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridises under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subsequence, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above). The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterising gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docID=20030194704
 XX

CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above). The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterising
 CC alternative splicing events, in detecting and characterising gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The present sequence for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 573 BP; 86 A; 202 C; 171 G; 114 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2.55e-37 Length: 573
 Score: 367.00 Matches: 70
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.07% Indels: 0
 DB: 12 Gaps: 0

US-10-657-740-1 (1-173) x ACH79686 (1-573)

Qy 104 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 123
 Db 203 CAGGACGACACCGGCTACATTCCGTCGAGTTCACCCGCGTACCGCTCCGCTCCCAAC 262
 Qy 124 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 143
 Db 263 GTGGACCAAGTCGGCCCTCTTGCTCCCTGTCGCGATGCGATGCTGACCTTCTGTGGC 322
 Qy 144 ProlysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 163
 Db 323 CCCAAGATCCAGACTGGCTGATGCCACCCACGCGGAGCGGCATCCCGTGTGCGGG 382
 Qy 164 GluGluLysProThrSerAlaProSerSer 173
 Db 383 GAGGAGAGCCACCTCGGCTCCCTCGTCC

RESULT 33
 ACN89210
 ID ACN89210 standard; DNA; 411 BP.
 XX
 AC ACN89210;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Breast cancer related marker, seq id 10360.
 XX
 KW Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN US2003099974-A1.
 XX
 PD 29-MAY-2003.
 XX
 PF 18-JUL-2002; 2002US-00198846.
 XX
 PR 18-JUL-2001; 2001US-0306220P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2003-787014/74.
 DR
 XX
 PT Novel isolated polypeptide associated with breast cancer, useful for
 PT detecting presence of polypeptide in sample, as a marker for breast
 PT cancer.
 XX
 PS Disclosure; SEQ ID NO 10360; 36pp; English.
 XX
 CC The invention relates to an isolated polypeptide (I) associated with
 CC breast cancer which is encoded by a nucleic acid molecule comprising a
 CC nucleotide sequence (S1). Further disclosed is an antibody that binds to
 CC the polypeptide of the invention. The activity of the polypeptide of the
 CC invention may be described as cytostatic. The antibody is useful for
 CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
 CC invention are useful in the detection of breast tumours. (I) is useful as
 CC a marker for breast cancer and in breast cancer therapy. Sequences given
 CC in records ACN78851-ACN92934 represent nucleic acid markers associated
 CC with breast cancer. Note: The sequence listing does not form part of the
 CC specification but may be obtained in electronic format from the USPTO web
 CC site at seqdata.uspto.gov/sequence.html?DocID=20030099974
 XX
 SQ Sequence 411 BP; 86 A; 134 C; 99 G; 92 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3.26e-37 Length: 411
 Score: 364.50 Matches: 71
 Percent Similarity: 75.40% Conservative: 24
 Best Local Similarity: 56.35% Mismatches: 24
 Query Match: 39.79% Indels: 7
 DB: 11 Gaps: 5

US-10-657-740-1 (1-173) x ACN89210 (1-411)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
 Db .39 ATGGACATAGCATCCACACCCCTGGATCGGGTCCCTCTTTTCTTCCACTCCCA 98
 Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
 Db 99 AGCGCGCTCTTTGACCAAGTCTTCGAGAGCACCTGTGGAGTCTGACCTCTCTCT--- 155
 Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
 Db 156 ACAGCCACTTCCCTGAGCCCTTCTACTTCGGCCACCCCTCTTCCGGGGCACCCTAGC 215
 Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
 Db 216 TGGATTGACACTGGGCTCTCAGAGATCGTATGGAAGGACAGGTTCTCTGTGAACCTG 275
 Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
 Db 276 GACGTGAAGCACTTCTCTCAGAGGAAGTCAAGGTCTGCGAGAGCGTATTGAG 335
 Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
 Db 336 GTGCACGGCAGACACGAGAGCGCCAGACATGGCTTCACTCTCCAGGAGGTTCCAC 395
 Qy 116 ArgArgTyrArgLeuPro 121
 Db 396 AGGAAGTAC---CTGCC 410

RESULT 34
 ACH17737
 ID ACH17737 standard; cDNA; 450 BP.
 XX
 AC ACH17737;
 XX
 DT 13-OCT-2003 (first entry)
 XX
 DE Human adult heart cDNA #2051.
 XX
 KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;

CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20030194704
 XX
 SQ Sequence 579 BP; 89 A; 200 C; 163 G; 127 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 9,84e-35 Length: 579
 Score: 347.00 Matches: 65
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 37.88% Indels: 0
 DB: 12 Gaps: 0

US-10-657-740-1 (1-173) x ACH67866 (1-579)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGluProPheTyrProSer 20
 Db 146 ATGGACGTGACCATCAGCACCCCTGGTTCAAGCGCACCTGGGGCCCTTCTACCCAGC 205
 QY 21 ArgLeuPheAspGlnPhePheGluGluGluLeuPheGluTyrAspLeuLeuProPheLeu 40
 Db 206 CGGCTGTTGACACAGTTTTCGGCGAGGGCCCTTTTGAGTATGACCTGCTGCCCTTCCTG 265
 QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 Db 266 TCGTCCACCATCAGCCCTTACTACCGCCAGTCTCCCTTCGCGACCGTGTGACCTCGGC 325
 QY 61 IleSerGluValArg 65
 Db 326 ATCTCTGAGGTAAGA 340

RESULT 36

ADH76210
 ID ADH76210 standard; DNA; 480 BP.
 XX
 AC ADH76210;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE Human heat shock protein 20 gene SEQ ID NO:303.
 XX
 KW heat shock protein 20; cytosolic; antiarteriosclerotic; vasotropic;
 KW antianginal; cerebroprotective; antiarrhythmic; antiasthmatic;
 KW synaerological; hypotensive; antimigraine; tocolytic; relaxant; HSP;
 KW smooth muscle cell; smooth muscle cell; ds; gene; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003018758-A2.
 XX
 PD 06-MAR-2003.
 XX
 PF 23-AUG-2002; 2002WO-US026918.
 XX
 PR 23-AUG-2001; 2001US-0314535P.
 XX

(UYAR-) UNIV ARIZONA.
 PA (BROP/) BROPHY C.
 PA (KOMA/) KOMALAVILAS P.
 PA (PANI/) PANITCH A.
 PA (SEAL/) SEAL B.

(LOKE/) LOKESH J.

PI Brophy C, Komalavilas P, Panitch A, Seal B, Lokesh J;

XX WPI; 2003-393248/37.

XX New heat shock protein 20-derived polypeptides, useful for inhibiting,
 PT treating or preventing smooth muscle cell vasospasm or a disorder such as
 PT intimal hyperplasia, restenosis, atherosclerosis or smooth muscle cell
 tumors.

XX Disclosure; SEQ ID NO 320; 194pp; English.

XX The invention relates to a novel polypeptide comprising a heat shock
 CC protein 20-derived polypeptide. A polypeptide of the invention has
 CC cytosolic, antiarteriosclerotic, vasotropic, antianginal,
 CC cerebroprotective, antiarrhythmic, antiasthmatic, synaerological,
 CC hypotensive, antimigraine, tocolytic, and relaxant activity, and may act
 CC as a HSP agonist or antagonist. The polypeptides, heat shock protein
 CC (HSP) 20, and methods are useful for treating or preventing a disorder,
 CC e.g. intimal hyperplasia, stenosis, restenosis, transplant vasculopathy,
 CC atherosclerosis, smooth muscle cell tumours such as leiomyosarcoma, or
 CC Prinzmetal's angina, coronary ischaemia, stroke, bradycardia,
 CC hypertension, pulmonary (lung) hypertension, asthma (bronchospasm),
 CC toxemia of pregnancy, pre-term labour, pre-eclampsia/eclampsia, Raynaud's
 CC disease, Raynaud's phenomenon, haemolytic-uremia, non-occlusive
 CC mesenteric ischaemia, anal fissure, achalasia, impotence, migraine, or
 CC ischaemic muscle injury associated with smooth muscle spasm. The
 CC polypeptides and HSP 20 are also useful for inhibiting smooth muscle cell
 CC proliferation and/or migration. The present sequence represents a peptide
 CC used in a polypeptide of the invention.

XX Sequence 480 BP; 110 A; 133 C; 131 G; 106 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5,95e-34 Length: 480
 Score: 340.00 Matches: 69
 Percent Similarity: 58.14% Conservative: 31
 Best Local Similarity: 40.12% Mismatches: 50
 Query Match: 37.12% Indels: 22
 DB: 10 Gaps: 3

US-10-657-740-1 (1-173) x ADH76210 (1-480)

QY 3 ValThrIleGlnHisProTrpPheLysArgThrLeuGluProPhe-----Tyr 18
 Db 13 GTTCCAGTCCAGCCTAGTTGGCTAAGAAGAGCTAGTGCCTTTCCCGGTTTGAGTGCC 72
 QY 19 ProSerArgLeuPheAspGlnPheGlyGluGluLeuPheGluTyrAspLeuPro 38
 Db 73 CCGCGGAGCTATTGATCAACGCTTTCGGAGGGGTTACTGAGGCTGAATTAGCAGCA 132
 QY 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
 Db 133 CTTTGTCCGACACACTCGGCCCTATTAC-----CTTAGAGCGCGCTGTGAGCC 183
 QY 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLys 78
 Db 184 TTACCAGTGCCTCAGGTACCACTACCCAGGCCACTTCTCCGTTTATTATAGCGTGAAA 243
 QY 79 HisPheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGly 98
 Db 244 CACTTTAGCCCAAGAGATGACAGTCAAGTTGTAGGAGCATGTGGAAGTTACGCG 303
 QY 99 LysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgTyr 118
 Db 304 AGACATGAAGAGAGACAGATGAACATGCTTTCGTAGCGAGAGAATTCATCGCGTAT 363
 QY 119 ArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
 Db 364 CGTCTGCCCCCAGGAGTGCATCTCGACGTGTGACGAGTGCAATTATCGCCTGAGGAGTG 423

Qy 139 LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisalaGluArgAla 150
Db 424 CTTAGTATC-----CAAGCA 438
Qy 159 IleProValSerArgGluGluLysProThrSerAla 170
Db 439 GCCCCGGCGTCAGCCCAAGCCCCCGCTCGCGTGCT 474

RESULT 37
ABQ61117
ID ABQ61117 standard; cDNA; 575 BP.
XX
AC ABQ61117;
XX
DT 26-FEB-2003 (first entry)
XX
DE Skeletal muscle stress protein p20 encoding sequence.
XX
KW Neuroprotective; immunomodulator; cancer; chromosome 19pter-q12;
KW cystostatic; anti-inflammatory; gene therapy; nutritional supplement;
KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnary; gene; ss.
XX
OS Homo sapiens.
XX
OS
XX
PN WO200231111-A2.
XX
PD 18-APR-2002.
XX
PF 11-OCT-2001; 2001WO-US027760.
XX
PR 12-OCT-2000; 2000US-00687527.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
PI
XX WPI; 2002-426278/45.
DR N-PSDB; ABP43873.
XX
XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.
XX
PS Claim 1; SEQ ID # 330; 357pp + Sequence Listing; English.
XX
CC The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnary, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABQ60788-
CC ABQ61233 represent polynucleotides of the invention. NOTE: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC fp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 575 BP; 87 A; 216 C; 183 G; 89 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 7,79e-34 Length: 575
Score: 340.00 Matches: 69
Percent Similarity: 58.14% Conservative: 31
Best Local Similarity: 40.12% Mismatches: 50
Query Match: 37.12% Indels: 22
DB: 6 Gaps: 3

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. The antibodies to the proteins can also be used
 CC in alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. cardiac
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cerebral
 CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
 CC The polypeptides can also be used to aid wound healing and epithelial
 CC cell proliferation, to prevent skin aging due to sunburn, to maintain
 CC organs before transplantation, for supporting cell culture of primary
 CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
 CC also be used as a food additive or preservative to increase or decrease
 CC storage capabilities. This sequence corresponds to a gene of the
 CC invention.

XX
 SQ Sequence 1433 BP; 260 A; 592 C; 307 G; 263 T; 0 U; 11 Other;

Alignment Scores:
 Pred. No.: 3,056-33 Length: 1433
 Score: 340.00 Matches: 69
 Percent Similarity: 58.14% Conservative: 31
 Best Local Similarity: 40.12% Mismatches: 50
 Query Match: 37.12% Indels: 22
 DB: 5 Gaps: 3

US-10-657-740-1 (1-173) x ADM19230 (1-1433)

Qy 3 ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProphe-----Tyr 18
 Db 43 GTGCTGTGAGCGCTTGTGCTGCGCGCGCTCGCGCGCTTCCCGGACTTTCGCG 102
 Qy 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
 Db 103 CCCGACGCTCTTTGACGAGCGCTTCGCGAGGGGCTGTGGAGCGCGAGCTGCTGCG 162
 Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
 Db 163 CTCTGCCGCCACCGCTCGCCCTCTACTAC-----CTGCGCGCACCCAGCGTGGCG 213
 Qy 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78
 Db 214 CTGCGCGTCCGCGAGGATGCTGTCAGGTGGTGGCGAACAACGTCGAGGTGCACGCG 273
 Qy 79 HisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIleHisGly 98
 Db 274 CACTTCTCGCGAGGAAATTGCTGTCAGGTGGTGGCGAACAACGTCGAGGTGCACGCG 333
 Qy 99 LysHisArgGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyr 118
 Db 334 CGCCACGAGGAGCGCGCGATGAGCACGATTCGTGCGCGCGAGTTCACCGCTGCTAC 393
 Qy 119 ArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
 Db 394 CGCTTCGCGCTCGGTGGATCCGCTCGGTGAGTCCGCTGCTGCTGCTGCTGCTGCTGCT 453
 Qy 139 LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAla 158
 Db 454 CTGTCCATC-----CAGGCC 468
 Qy 159 IleProValSerArgGluGluLysProThrSerAla 170
 Db 469 GCACGAGCTGCA 504
 RESULT 39
 AAZ42230
 ID AAZ42230 standard; cDNA; 1488 BP.
 XX
 AC AAZ42230;

XX 31-JAN-2000 (first entry)
 DT Human normal bladder tissue cDNA derived EST 109.
 DE Human; bladder; treatment; EST; expressed sequence tag; cytostatic;
 KW cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 PN DE19818620-A1.
 XX 28-OCT-1999.
 PF 21-APR-1998; 98DE-01018620.
 XX 21-APR-1998; 98DE-01018620.
 XX (METR-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 FI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
 XX
 DR WPI; 1999-602416/52.
 DR P-PSDB; AAY60544, AAY60545, AAY60546.
 PT New polypeptides and their nucleic acids, useful for treatment of bladder
 PT tumor and identification of therapeutic agents.
 PS Claim 3; Page 230; 366pp; German.
 XX

CC This invention describes novel polypeptide fragment sequences (I) and
 CC their encoding nucleic acids (II) which are highly expressed in normal
 CC bladder tissue and have cytostatic activity. (II) are used for
 CC recombinant expression of (I) and to isolate complete genes. (I) are used
 CC to identify agents suitable for the treatment of bladder tumours, to
 CC directly treat this form of cancer (including expression from gene
 CC therapy vectors), or are used in a preparation for cancer treatment. (I)
 CC is also used for the generation of specific antibodies. (II) are
 CC identified by assembling ESTs (expressed sequence tags) from a particular
 CC tissue type before comparison of expression patterns. This allows a
 CC significantly longer fragment of the gene to be revealed, and therefore
 CC reduces the number of failures because of ESTs from different libraries
 CC representing different parts of the same unknown gene distorting the
 CC estimated frequency of occurrence in a particular tissue. AAZ42122-242248
 CC represent EST fragments derived from a human normal bladder tissue cDNA
 CC library which encode the protein fragments represented in AAY60329-Y60591
 XX
 SQ Sequence 1488 BP; 287 A; 607 C; 321 G; 273 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,236-33 Length: 1488
 Score: 340.00 Matches: 69
 Percent Similarity: 58.14% Conservative: 31
 Best Local Similarity: 40.12% Mismatches: 50
 Query Match: 37.12% Indels: 22
 DB: 2 Gaps: 3

US-10-657-740-1 (1-173) x AAZ42230 (1-1488)

Qy 3 ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProphe-----Tyr 18
 Db 29 GTGCTGTGACGCGCTTGTGCTGCGCGCGCTTCCCGGACTTTCGCGCG 88
 Qy 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
 Db 89 CCGGAGCGCTCTTTGACGAGCGCTTCGCGAGGGGCTGTGGAGCGCGAGCTGCTGCG 148
 Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
 Db 149 CTCTGCCGCCACCGCTCGCCCTCTACTAC-----CTGCGCGCACCCAGCGTGGCG 199
 Qy 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78
 Db 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78

```
Db      200  CTGCCCGTGGCCAGGTGCCGACGACCCCGCCACATTTTCGGTGCTGCTAGACGTGAAG 259
Qy      79  HisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGluIleHisGly 98
Db      260  CACTTCTCGCGGAGAAATTGCTGAAGTGTGGCGAACACAGTGGAGGTGCACGCG 319
Qy      99  LysHisAenGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118
Db      320  CGCCACGAGGAGCGCCCGATGAGACGATTCGTGGCGCGGAGTTCACCGTCGCTAC 379
Qy      119  ArgLeuProSerAenValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
Db      380  CGCCTGCCCGCTGGCGTGGATCGCGCTCCGACGTCGCGCTGCCCGGAGGCGTC 439
Qy      139  LeuThrPheCysGlyProIysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAla 158
Db      440  CTGTCCATC-----CAGGCC 454
Qy      159  IleProValSerArgGluGluLysProThrSerAla 170
Db      455  GCACCAGCGTCGCCCGCCAGGCCCGCCCGCCAGCGCA 490
RESULT 40
ADM19501/c
ID      ADM19501 standard; cDNA; 1506 BP.
XX
AC      ADM19501;
XX
XX      20-MAY-2004 (first entry)
XX
XX      Novel human channel/transporter gene #27 clone 2.
XX
KW      ds; gene; immunosuppressive; antiarthritic; antirheumatic;
KW      antiproliferative; cytostatic; cardiant; vasotrophic; cerebroprotective;
KW      nootropic; neuroprotective; antibacterial; virucide; fungicide;
KW      ophthalmological; gene therapy; channel/transporter protein;
KW      rheumatoid arthritis; neoplasm; cardiac arrest; cerebrovascular disorder;
KW      cerebral ischemia; angiogenesis; nervous system disorder;
KW      Alzheimer's disease; ocular disorder; corneal infection; wound healing;
KW      epithelial cell proliferation; skin aging; sunburn; transplantation;
KW      chemotaxis; food additive.
XX
OS      Homo sapiens.
XX
XX      WO200154472-A2.
XX
XX      02-AUG-2001.
XX
XX      17-JAN-2001; 2001WO-US001307.
XX
XX      31-JAN-2000; 2000US-0179065P.
FR      04-FEB-2000; 2000US-0180628P.
FR      24-FEB-2000; 2000US-0184664P.
PR      02-MAR-2000; 2000US-0186350P.
PR      16-MAR-2000; 2000US-0189874P.
PR      17-MAR-2000; 2000US-0190076P.
PR      18-APR-2000; 2000US-0198123P.
PR      19-MAY-2000; 2000US-0205515P.
PR      07-JUN-2000; 2000US-0209467P.
PR      28-JUN-2000; 2000US-0214886P.
PR      30-JUN-2000; 2000US-0215135P.
PR      07-JUL-2000; 2000US-0216647P.
PR      07-JUL-2000; 2000US-0216880P.
PR      11-JUL-2000; 2000US-0217487P.
PR      11-JUL-2000; 2000US-0217496P.
PR      14-JUL-2000; 2000US-0218290P.
PR      26-JUL-2000; 2000US-0220963P.
PR      26-JUL-2000; 2000US-0220964P.
PR      14-AUG-2000; 2000US-0224518P.
PR      14-AUG-2000; 2000US-0224519P.
PR      14-AUG-2000; 2000US-0225213P.
PR      14-AUG-2000; 2000US-0225214P.
PR      14-AUG-2000; 2000US-0225266P.
PR      14-AUG-2000; 2000US-0225267P.
PR      14-AUG-2000; 2000US-0225268P.
PR      14-AUG-2000; 2000US-0225270P.
PR      14-AUG-2000; 2000US-0225447P.
PR      14-AUG-2000; 2000US-0225757P.
PR      14-AUG-2000; 2000US-0225758P.
PR      14-AUG-2000; 2000US-0225759P.
PR      18-AUG-2000; 2000US-0226279P.
PR      22-AUG-2000; 2000US-0226681P.
PR      22-AUG-2000; 2000US-0226868P.
PR      23-AUG-2000; 2000US-0227182P.
PR      30-AUG-2000; 2000US-0228924P.
PR      01-SEP-2000; 2000US-0229287P.
PR      01-SEP-2000; 2000US-0229343P.
PR      01-SEP-2000; 2000US-0229344P.
PR      05-SEP-2000; 2000US-0229345P.
PR      05-SEP-2000; 2000US-0229509P.
PR      06-SEP-2000; 2000US-0230437P.
PR      06-SEP-2000; 2000US-0230438P.
PR      08-SEP-2000; 2000US-0231242P.
PR      08-SEP-2000; 2000US-0231243P.
PR      08-SEP-2000; 2000US-0231244P.
PR      08-SEP-2000; 2000US-0231413P.
PR      08-SEP-2000; 2000US-0231414P.
PR      08-SEP-2000; 2000US-0232080P.
PR      08-SEP-2000; 2000US-0232081P.
PR      12-SEP-2000; 2000US-0231968P.
PR      14-SEP-2000; 2000US-0232397P.
PR      14-SEP-2000; 2000US-0232398P.
PR      14-SEP-2000; 2000US-0232399P.
PR      14-SEP-2000; 2000US-0232400P.
PR      14-SEP-2000; 2000US-0232401P.
PR      14-SEP-2000; 2000US-0233063P.
PR      14-SEP-2000; 2000US-0233064P.
PR      14-SEP-2000; 2000US-0233065P.
PR      21-SEP-2000; 2000US-0234223P.
PR      21-SEP-2000; 2000US-0234274P.
PR      25-SEP-2000; 2000US-0234997P.
PR      25-SEP-2000; 2000US-0234998P.
PR      26-SEP-2000; 2000US-0235484P.
PR      27-SEP-2000; 2000US-0235836P.
PR      29-SEP-2000; 2000US-0236327P.
PR      29-SEP-2000; 2000US-0236367P.
PR      29-SEP-2000; 2000US-0236368P.
PR      29-SEP-2000; 2000US-0236369P.
PR      29-SEP-2000; 2000US-0236370P.
PR      02-OCT-2000; 2000US-0236802P.
PR      02-OCT-2000; 2000US-0237037P.
PR      02-OCT-2000; 2000US-0237038P.
PR      02-OCT-2000; 2000US-0237039P.
PR      02-OCT-2000; 2000US-0237040P.
PR      13-OCT-2000; 2000US-0239935P.
PR      13-OCT-2000; 2000US-0239937P.
PR      20-OCT-2000; 2000US-0240960P.
PR      20-OCT-2000; 2000US-0241212P.
PR      20-OCT-2000; 2000US-0241785P.
PR      20-OCT-2000; 2000US-0241786P.
PR      20-OCT-2000; 2000US-0241787P.
PR      20-OCT-2000; 2000US-0241808P.
PR      20-OCT-2000; 2000US-0241809P.
PR      20-OCT-2000; 2000US-0241826P.
PR      01-NOV-2000; 2000US-0244617P.
PR      08-NOV-2000; 2000US-0246474P.
PR      08-NOV-2000; 2000US-0246475P.
PR      08-NOV-2000; 2000US-0246476P.
PR      08-NOV-2000; 2000US-0246477P.
PR      08-NOV-2000; 2000US-0246478P.
PR      08-NOV-2000; 2000US-0246523P.
PR      08-NOV-2000; 2000US-0246524P.
PR      08-NOV-2000; 2000US-0246525P.
```

PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX
 DR
 DR
 DR
 XX
 PT
 XX
 PS

WPI; 2001-476159/51.
 P-PSDB; ADM19980.

Isolated nucleic acid molecule encoding a channel/transporter protein is used in preventing, treating or ameliorating a medical condition.

Claim 1; SEQ ID NO 308; 809pp; English.

The invention relates to an isolated nucleic acid molecule encoding a channel/transporter protein or sequences at least 95% identical to a these. The nucleic acids and proteins encoded by them are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease

CC storage capabilities. This sequence corresponds to a gene of the invention.
 XX

SQ Sequence 1506 BP; 276 A; 318 C; 616 G; 295 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 3.28e-33 Length: 1506
 Score: 340.00 Matches: 69
 Percent Similarity: 58.14% Conservatives: 31
 Best Local Similarity: 40.12% Mismatches: 50
 Query Match: 37.12% Indels: 22
 DB: 5 Gaps: 3

US-10-657-740-1 (1-173) x ADM19501 (1-1506)

Qy 3 ValThrIleGlnHisProTyrPheLeuValArgThrLeuGlyProPhe-----Tyr 18
 Db 1475 GTGCTGTGCGCGCTTGTGGCTGGCGCGCGCTTGGCGCGCGCTTGGCGCGCGCTTGGCGG 1416
 Qy 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
 Db 1415 CCCGAGCGCTCTTTGACACGCTTGGCGAGGGGCTGTGGAGGCGCGAGCTGGCTGCG 1356
 Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
 Db 1355 CTGTGCCCCACCGCTGCGCCCTACTAC-----CTGCGCGCACCCAGCGTGGCG 1305
 Qy 59 SerGlyIleSerGluValArgSerArgAspIysPheValIlePheLeuAspVallys 78
 Db 1304 CTGCGCGTGGCGCGAGTGGCGACCGACCGCGCGCGCTTTCGGTGTCTAGAGTGAAG 1245
 Qy 79 HisPheSerProGluAspLeuThrValValysValGlnAspAspPheValGluIleHisGly 98
 Db 1244 CACTTCTCGCGGAGGAATTGCTCAAGTGTGGCGGAACACGTGGAGGTGCACGCG 1185
 Qy 99 LysHisGlnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118
 Db 1184 CGCCACGAGGAGCGCGCGATGACACGAGATTCGTGCGCGCGAGTTCCACCGTGGCTAC 1125
 Qy 119 ArgLeuProSerAanValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
 Db 1124 CGCTGCGCGCTGGCTGGATTCGGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 1065
 Qy 139 LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluAla 158
 Db 1064 CTGTCCATC-----CAGGCC 1050
 Qy 159 IleProValSerArgGluGluLysProThrSerAla 170
 Db 1049 GCACCGAGCTGCGCGCCAGGCCCGCCAGGCCCGCGCGCA 1014

RESULT 41
 AAC76681

ID AAC76681 standard; cDNA; 1771 BP.

AC AAC76681;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORP2236 polynucleotide sequence SEQ ID NO:4471.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 XX 05-OCT-2000.
 XX
 XX 31-MAR-2000; 2000WO-US008621.
 XX
 XX 31-MAR-1999; 99US-0127607P.
 PR 02-APR-1999; 99US-0127636P.
 PR 05-APR-1999; 99US-0127728P.
 PR 30-MAR-2000; 2000US-00540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX
 XX Shimkets RA, Leach M;
 XX
 XX WPI; 2000-602362/57.
 DR P-PSDB; AAB42472.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease.
 XX
 XX Claim 5; Page 3656-3657; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antipsoriatic; antiparinsonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiact; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
 CC antiviral; antifungal; antirheumatic; antichyroid; and antianaemic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to or preventing or treating pathological conditions associated with an
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
 XX
 XX Sequence 1771 BP; 342 A; 704 C; 389 G; 336 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 4.18e-33 Length: 1771
 Score: 340.00 Matches: 69
 Percent Similarity: 58.14% Conservative: 31
 Best Local Similarity: 40.12% Mismatches: 50
 Query Match: 37.12% Indels: 22
 DB: 3 Gaps: 3
 US-10-657-740-1 (1-173) x AAC76681 (1-1771)
 Qy 3 ValThrIleGlnHisProTyrPheLeuGlyProPhe-----Tyr 18
 Db 313 GTGCTGTGACGCGTCTGTGCTGCGCGCCCTCGCGCCGTTGCCGGACTTTCGCGC 372
 Qy 19 ProSerArgLeuPheAspGlnPheGlyGluLeuPheGluTyrAspLeuLeuPro 38
 Db 373 CCGGACGCCCTCTTGACACGCGTTGCGGAGGGGCTGTGGAGCGGAGCTGGCTGCG 432
 Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
 Db 433 CTCTGCCCCACGCTCGCCCTACTACTAC-----CTGCGCGCACCCGCGTGGCG 483

Qy 59 SerGlyIleSerGluValArgSerArgAspLysPheValIlePheLeuAspVallys 78
 Db 484 CTGCGCGTGCAGGTCGCGACGACCCCGCCACCTTTTCGGTGTCTAGAGCTGAAG 543
 Qy 79 HisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGly 98
 Db 544 CACTTCTCGCGGAGGAAATTGTGTCAAGGTGTGGCGAACAACGTGGAGGTGCACGCG 603
 Qy 99 LysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118
 Db 604 CGCCACGAGGAGGCCCGGATGACACGAGATTCGTCGCGCGAGTTCACCGCTGCTAC 663
 Qy 119 ArgLeuProSerAenValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
 Db 664 CGCTGCGCGCTGGCGTGGATCCGCTGCGTACGTCGCGCTGTCCCCGAGGGGCTC 723
 Qy 139 LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAla 158
 Db 724 CTGTCCATC-----CAGGCC 738
 Qy 159 IleProValSerArgGluGluLysProThrSerAla 170
 Db 739 GCACACGCTCGCGCCAGGCGCCACCCGCGCAGCGCA 774
 RESULT 42
 ADQ22255 ID ADQ22255 standard; DNA; 1820 BP.
 XX AC ADQ22255;
 XX 26-AUG-2004 (first entry)
 DT Human soft tissue sarcoma-upregulated DNA - SEQ ID 5075.
 DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 XX ds.
 XX Homo sapiens.
 OS WO2004048938-A2.
 FN 10-JUN-2004.
 PD 26-NOV-2003; 2003WO-US038193.
 PF 26-NOV-2002; 2002US-0429739P.
 PR (PROT-) PROTEIN DESIGN LABS INC.
 XX
 XX Pi Ariz N, Ginsburg WM, Zlotnik A;
 XX WPI; 2004-441208/41.
 DR
 XX Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX
 XX Example 2; SEQ ID NO 5075; 210pp; English.
 PS
 XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytosatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the

```
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 1820 BP; 374 A; 822 C; 335 G; 289 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4,366-33 Length: 1820
Score: 340.00 Matches: 69
Percent Similarity: 58.14% Conservative: 31
Best Local Similarity: 40.12% Mismatches: 50
Query Match: 37.12% Indels: 22
DB: 12 Gaps: 3

US-10-657-740-1 (1-173) x AD022255 (1-1820)

Qy 3 ValThrIleGlnHisProTyrPheLeuArgThrLeuGlyProPhe-----Tyr 18
Db 35 GTGCGTGTGCAGCCGCTTGTGGTGTGCGCGCGCTCGGCCCGCTTCCCGGACTTTCGGCG 94

Qy 19 ProSerArgLeuPheAspGlnPheGlyGluGlyLeuPheGlyTyrAspLeuLeuPro 38
Db 95 CCGGACGCCCTTTGACAGCGCTTCGGCGAGGGGCTGCTGGAGCGGAGCTGGCTGCG 154

Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
Db 155 CTCTGCCCCACACGCTCGCCCTCTACTAC-----CTGGCGGCACCCAGCGTGGCG 205

Qy 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78
Db 206 CTGCCCGTGGCCAGTCCGAGCGAGCGCCCGCCACTTTTCGGTGTCTAGAGTGAAG 265

Qy 79 HisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGluIleHisGly 98
Db 266 CACTTCTCGCGGAGGAATTGCTGCAAGTGTGTGGCGCAACACAGTGGAGGTGCACGCG 325

Qy 99 LysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118
Db 326 CGCACAGAGAGCGCCGAGTGTGAGCAGCGATTCGTGGCGCGGAGTTCACCGTGCCTAC 385

Qy 119 ArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
Db 386 CGCTGCGCGCTGGCGTGGATCCGCTGCGTGCAGCTCCGCTGTCCTCCCGAGGGGCTC 445

Qy 139 LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAla 158
Db 446 CTGTCCATC-----CAGGCC 460

Qy 159 IleProValSerArgGluGluLysProThrSerAla 170
Db 461 GCACAGCGTGGCCCGAGCCCGCCCGCCAGCCGCA 496

RESULT 43
ACH81569
ID ACH81569 standard; DNA; 212 BP.
XX
AC ACH81569;
XX
XX 29-JUL-2004 (first entry)
XX
XX Human genome derived single exon probe #14764.
XX
XX Human; probe; ss; gene expression; single exon probe; microarray;
XX alternative splicing event; genomic alteration.
XX
XX Homo sapiens.
XX
XX US2003194704-A1.
XX
XX 16-OCT-2003.
XX
XX 03-APR-2002; 2002US-00029386.
XX
XX 03-APR-2002; 2002US-00029386.
XX
```

(PENN/) PENN S G.
(RANK/) RANK D R.
(HANZ/) HANZEL D K.

Penn SG, Rank DR, Hanzel DK;

WPI; 2004-119264/12.

New human genome-derived single exon nucleic acid probes useful for human gene expression analysis, for identifying or characterizing alternative splicing events, for assessing genomic alterations or as tools for surveying tissues.

Claim 1; SEQ ID NO 14764; 80pp; English.

The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 6888 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridises under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human gene expression (comprising a plurality of single exon nucleic acid probes cited above, where each of the plurality of probes is separately and addressably isolatable or amplifiable from the plurality), a single exon microarray for measuring human gene expression, a method of measuring human gene expression, a vector comprising the single exon probe cited above, an ORF-encoded peptide comprising at least 8 contiguous amino acids of any of the above-mentioned amino acid sequences (optionally with conservative amino acid substitutions), an isolated antibody that binds specifically to a peptide cited above, methods of selling and/or licensing single exon probes or microarrays to a customer desiring to measure gene expression, a method of providing human gene expression data by subsequence, and a computer-readable storage medium which contains a database having a plurality of records (each record including data on the expression of a single exon probe cited above). The probe, methods and apparatus are useful in gene expression analysis. The probes may be used as tools for surveying tissues to detect the presence of expressed messages that contain their specific exon, or in constructing genome-derived single exon microarrays. In addition, the probes are used in identifying and characterising alternative splicing events, in detecting and characterising gross alterations in the genomic locus that includes their exon, in assessing smaller genomic alterations, in priming the synthesis of nucleic acids, or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030194704

Sequence 212 BP; 33 A; 81 C; 49 G; 49 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,186-34 Length: 212
Score: 338.00 Matches: 63
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.90% Indels: 0
DB: 12 Gaps: 0

US-10-657-740-1 (1-173) x ACH81569 (1-212)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 24 ATGGACGTGACCATCCAGCACCCCTGTTTCAGCGCACCCCTGGGGGCCCTTACCCCGC 83
Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGlyTyrAspLeuLeuProPheLeu 40
Db 84 CGGCTGTTCGACCAAGTTTTCGGCGAGGGCTTTTCAGTATGACCTGCTGCTCTCTG 143
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60

Db 144 TGTCCACATCAGCCCTACTACCGCCAGTCCCTCTTCCGACCGTGTGGACTCCGCGC 203

Qy 61 IleSerGlu 63

Db 204 ATCTGTAG 212

RESULT 44

ACH17560

ID ACH17560 standard; cDNA; 421 BP.

XX ACH17560;

XX 13-OCT-2003 (first entry)

XX Human adult heart cDNA #1874.

XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.

XX Homo sapiens.

XX US2003073623-A1.

XX 17-APR-2003.

XX 30-JUL-2001; 2001US-00918995.

XX 30-JUL-2001; 2001US-00918995.

XX (DRMA/) DRMANAC R T.

XX (LABA/) LABAT I.

XX (STAC/) STACHE-CRAIN B.

XX (DICK/) DICKSON M C.

XX (JONE/) JONES L W.

XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;

XX WPI; 2003-615964/58.

XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.

PS Claim 1; SEQ ID NO 4772; 44pp; English.

XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversity, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?docID=20030073623

XX Sequence 421 BP; 90 A; 133 C; 94 G; 104 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	8.86e-34	Length:	421
Score:	338.00	Matches:	68
Percent Similarity:	72.26%	Conservative:	31
Best Local Similarity:	49.64%	Mismatches:	32
Query Match:	36.90%	Indels:	7
DB:	9	Gaps:	4

US-10-657-740-1 (1-173) x ACH17560 (1-421)

Qy 1 MetAspValThrIleGlnHisProTyrPheIleValArgThrLeuGlyProPheTyrPro--- 19

Db 15 ATGACATCGCCATCCACCCCTTGGATCGCGGCCCTTCTTTCTTTCATCTCCGCGC 74

Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProphe 39

Db 75 GGCGCCTCTTAGACCATTTCTTCGGAGAGACCTGTTGGAGTCTGATCTTTTCCCG--- 131

Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55

Db 132 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGCCACACCTCTTCTTCGCGGCACCCAGC 191

Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLeuValIlePheLeu 75

Db 192 TGGGTTGACACTGGACTCTCAAGATCGCCTGAGAGGACAGATCTCTGTCAACCTG 251

Qy 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspAspPheValGlu 95

Db 252 GATGTGAGGACTTCTCCCAAGGAACCTCAAGGTAAAGTGTG-GGAGAGCTGATTGAC 310

Qy 96 IleHisGlyLysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115

Db 311 GTGCATGGTAAACATGACGAGCGCAGATGAACATGGTTTCATCTCCAGGGAGTTCCAC 370

Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSer 132

Db 371 AGGAATACCGGATCCAGCTGATGATAGACCTCTTACCATTACTTCATCC 421

RESULT 45

ADH76202

ID ADH76202 standard; DNA; 486 BP.

XX ADH76202;

XX 22-APR-2004 (first entry)

DE Rat heat shock protein 20 gene SEQ ID NO:303.

XX heat shock protein 20; cytostatic; antiarteriosclerotic; vasotropic;

XX antiangiinal; cerebroprotective; antiarrhythmic; antiasthmatic;

XX gynaecological; hypotensive; antimigraine; tocolytic; relaxant; HSP;

XX smooth muscle cell; smooth muscle cell; ds; gene.

XX Rattus sp.

XX WO2003018758-A2.

XX 06-MAR-2003.

XX 23-AUG-2002; 2002WO-US026918.

XX 23-AUG-2001; 2001US-0314535P.

XX (UVAR-) UNIV ARIZONA.

XX (BROP/) BROPHY C.

XX (KOMA/) KOMALAVILAS P.

XX (PANI/) PANITCH A.

XX (SEAL/) SEAL B.

XX (LOKE/) LOKESH J.

XX Brophy C, Komalavilas P, Panitch A, Seal B, Lokesh J;

XX WPI; 2003-393248/37.

XX New heat shock protein 20-derived polypeptides, useful for inhibiting,
PT treating or preventing smooth muscle cell vasospasm or a disorder such as
PT incinal hyperplasia, restenosis, atherosclerosis or smooth muscle cell
PT tumors.

XX Disclosure; SEQ ID NO 303; 194pp; English.

XX The invention relates to a novel polypeptide comprising a heat shock
CC protein 20-derived polypeptide. A polypeptide of the invention has
CC cytosolic, antiarteriosclerotic, vasotrophic, antianginal,
CC cerebroprotective, antiarrhythmic, antiasthmatic, gynaecological,
CC hypotensive, antimigraine, tocolytic, and relaxant activity, and may act
CC as a HSP agonist or antagonist. The polypeptides, heat shock protein
CC (HSP) 20, and methods are useful for treating or preventing a disorder,
CC e.g. intimal hyperplasia, stenosis, restenosis, transplant vasculopathy,
CC atherosclerosis, smooth muscle cell tumours such as leiomyosarcoma, or
CC Prinzmetal's angina, coronary ischaemia, stroke, bradycardia,
CC hypertension, pulmonary (lung) hypertension, asthma (bronchospasm),
CC toxemia of pregnancy, pre-term labour, pre-eclampsia/eclampsia, Raynaud's
CC disease, Raynaud's phenomenon, haemolytic-uremia, non-occlusive
CC mesenteric ischaemia, anal fissure, achalasia, impotence, migraine, or
CC ischaemic muscle injury associated with smooth muscle spasm. The
CC polypeptides and HSP 20 are also useful for inhibiting smooth muscle cell
CC proliferation and/or migration. The present sequence represents a peptide
CC used in a polypeptide of the invention.
XX
SQ Sequence 486 BP; 109 A; 139 C; 128 G; 110 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.99e-33	Length:	486
Score:	336.00	Matches:	71
Percent Similarity:	57.14%	Conservative:	29
Best Local Similarity:	40.57%	Mismatches:	53
Query Match:	36.68%	Indels:	22
DB:	10	Gaps:	3

US-10-657-740-1 (1-173) x ADH76202 (1-486)

Qy	3	ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPhe-----Tyr 18
Db	13	GTGCGGTACAAACCCAGCTGCTGGCGTCTCCGGCCATTACCTGGCTTACGTACC 72
Qy	19	ProSerArgLeuPheAspGlnPheGlyGluLeuPheGluTyrAspLeuPro 38
Db	73	CCCGACGATGTTTGGACGAGGTTTGGGAAGTTTACTTGAGCGGAATTAGCAAGT 132
Qy	39	PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
Db	133	CTATGTCTCGAGCTATAGTACCCCTACTAC-----CTAAGCGGCACCATCTCTCGCG 183
Qy	59	SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78
Db	184	CTCCCAACTGCCAAGTCCCCAGCATCCAGGCTATTTCAGCGTTCTGTTAGACGTAAAG 243
Qy	79	HisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGly 98
Db	244	CATTTAGTCCAGAAGAAATTCAGTAAAGTAGTGGGAGACCATGTCCAGGTACATGCT 303
Qy	99	LysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyr 118
Db	304	AGACACGAAGACAGACCTGATGAACACCGTTTCATCGTCGAGAGTTTCACCGCGTTAT 363
Qy	119	ArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
Db	364	CGCTTGCGCGCGGGGTTGATCCCGCGCGGCTCATCAGCACTCAGTCCGAGGAGGTT 423
Qy	139	LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAla 158
Db	424	TATCCATA-----CNAGCC 438
Qy	159	IleProValSerArgGluGluLysProThrSerAlaProSerSer 173
Db	439	ACACCGGCTCTGCTCAGCGCTCGTTCATCGCTCGCTCGCGCA 483

RESULT 46

ABT41710

ID ABT41710 standard; DNA; 1310 BP.

XX

AC ABT41710;

XX 26-JUN-2003 (first entry)

XX Toxicity modelling related rat gene SEQ ID No 1412.

XX Toxic effect; gene expression profile; renal toxicity; toxicity marker;
KW database; drug screening; toxicity assay; rat; ds.

XX Rattus norvegicus.

XX WO200295000-A2.

XX 28-NOV-2002.

XX 22-MAY-2002; 2002WO-US016173.

XX 22-MAY-2001; 2001US-0292335P.

XX 13-JUN-2001; 2001US-0297523P.

XX 19-JUN-2001; 2001US-0298925P.

XX 10-JUL-2001; 2001US-0303807P.

XX 10-JUL-2001; 2001US-0303808P.

XX 10-JUL-2001; 2001US-0303810P.

XX 28-AUG-2001; 2001US-0315047P.

XX 27-SEP-2001; 2001US-0324928P.

XX 22-OCT-2001; 2001US-0330462P.

XX 01-NOV-2001; 2001US-0330867P.

XX 26-DEC-2001; 2001US-0336144P.

XX 19-DEC-2001; 2001US-0340873P.

XX 21-FEB-2002; 2002US-0357842P.

XX 21-FEB-2002; 2002US-0357843P.

XX 15-MAR-2002; 2002US-0364134P.

XX 08-APR-2002; 2002US-0370144P.

XX 08-APR-2002; 2002US-0370206P.

XX 17-APR-2002; 2002US-0370247P.

XX 21-APR-2002; 2002US-0371679P.

(GENE-) GENE LOGIC INC.

Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

WPI; 2003-148464/14.

Predicting at least one toxic effect of a compound, useful for toxicity
modeling, comprises preparing a gene expression profile of a tissue or
cell sample exposed to the compound, and comparing the gene expression
profile to a database.

Example 4; Page; 446pp; English.

The invention relates to a novel method of predicting at least one toxic
effect of a compound. The method comprises a gene expression profile of a
tissue or cell sample exposed to the compound, and comparing the gene
expression profile to a database comprising at least part of the data or
information given in the specification. The methods are useful for
predicting at least one toxic effect of a compound, predicting the
progression of a toxic effect of a compound, predicting the renal
toxicity of a compound, or identifying toxicity markers in tissues or
cells exposed to known renal toxin. The genes are useful as toxicity
markers in drug screening and toxicity assays, in monitoring disease or
physiological status, or disease progression. This polynucleotide
represents a rat DNA sequence relating to the toxic effect database
described in the specification. NOTE: The sequence data for this patent
did not form part of the printed specification, but was obtained in
electronic format directly from the World Intellectual Property
Organization

SQ Sequence 1310 BP; 274 A; 446 C; 280 G; 310 T; 0 U; 0 Other;

Alignment Scores:


```

XX Sequence 405 BP; 50 A; 148 C; 137 G; 70 T; 0 U; 0 Other;
SQ Alignment Scores:
Pred. No.: 2,198-30 Length: 405
Score: 311.50 Matches: 60
Percent Similarity: 63.64% Conservative: 24
Best Local Similarity: 45.45% Mismatches: 41
Query Match: 34.01% Indels: 7
DB: 9 Gaps: 2

US-10-657-740-1 (1-173) x ABL07843 (1-405)
Qy 3 ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPhe-----Tyr 18
Db 16 GTGCTGTGCACCGCTTGTGGTGGCGCGCTCGGCCCTTCCCGAGACTTCGGCG 75

Qy 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
Db 76 CCGGACGCGCTCTTTGACACAGCGCTTCGGGAGGGGCTGCTGGAGGCGGAGCTGGCTGG 135

Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
Db 136 CTCTGCCCCACACCGCTCGCCCTTACTAC-----CTGCGCGCACCCAGCGTGGCG 186

Qy 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78
Db 187 CTGCGCGTGGCGGAGTGGCGGAGGACCCCGGCCACTTTTGGTCTGCTAGACGTGAAG 246

Qy 79 HisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIleHisGly 98
Db 247 CACTTCTCGCGGAGAAATTGCTGCTCAAGGTGGTGGCGAACAACGCTGGAGGTGCACGCG 306

Qy 99 LysHisAsnGluValArgGlnAspHisGlyTyrIleSerArgGluPheHisArgThrTyr 118
Db 307 CCACCAAGGAGGCGCGCGATGAGCAGCGATTCTGTCGCGCGCGAGTTCCACCGTGGCTAC 366

Qy 119 ArgLeuProSerAsnValAspGlnSerAlaLeuSer 130
Db 367 CCCTTGGCGCGCTGGTGTGATCCGCTGCGGTGACG 402

RESULT 50
ID ABL07843
XX ABL07843 standard; cDNA; 719 BP.
XX AC ABL07843;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18011.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX FA (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR P-PSDB; ABB63740.
XX New isolated nucleic acid detection reagent for detecting 1000 or more

```

```

PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
PS Claim'1; SEQ ID NO 18011; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 719 BP; 174 A; 208 C; 179 G; 158 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 1,26e-29 Length: 719
Score: 308.50 Matches: 69
Percent Similarity: 58.05% Conservative: 32
Best Local Similarity: 39.66% Mismatches: 54
Query Match: 33.68% Indels: 19
DB: 4 Gaps: 5

US-10-657-740-1 (1-173) x ABL07843 (1-719)
Qy 9 TrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPheGly 28
Db 77 TGGTGGGATGAGTCGACTTTCCAAATCGCACCTCGCGCTCTTGGACGACACTTCGGT 136

Qy 29 GluGlyLeuPheGluTyrAspLeuLeuProPheLeu----- 40
Db 137 CAGGAGCTGAAGAGGGATGACCTCATGCTCTCGGTGGAACACTCCGCCCCACAGTGTCTG 196

Qy 41 ---SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSer 59
Db 197 CGATCGGGATACCTGCGGCCCTTGGCACACGACCTGCGAGAGCAG-----GAGTCC 250

Qy 60 GlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHis 79
Db 251 GGT---TCCACCCCTCAACATCGACAGCGAGAAGTTCGAGGTCAATCTGGATGTGCAGCAG 307

Qy 80 PheSerProGluAspLeuThrVallysValGlnAspPheValGluIleHisGlyLys 99
Db 308 TTCTCGCCATCCGAGATCACCGTCAAGGTGGCCGCAAGTTCGTCATCGTGAGGGAAG 367

Qy 100 HisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArg 119
Db 368 CATGAGGAGAAGCAGACGACGAGATGCTCTCCGCCAGTTCTCCAGACGATTACCAG 427

Qy 120 LeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetIleu 139
Db 428 CTGCCAAGTGATGTCAACCCCGCACCGGTCAACATCGCTCCCTCTCTCCGATGGCTGCTG 487

Qy 140 ThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIle 159
Db 488 ACCATCAAGCGCCA-----ATGAAGCGCTGCTCTCTCCGACGACTGAACGCTTGTG 541

Qy 160 ProValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 542 CAGATCACAG-----ACTGGTCCCTCTCTCC 568

```

Search completed: May 30, 2005, 05:14:43
Job time : 550.598 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2005, 04:24:56 ; Search time 170.078 Seconds

(without alignments)
1664.391 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVTIQHPWFKRTGLPPFYS.....HAERAIPVSRREKPTSPSS 173

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10657740/runat_27052005_165254_3343/app_query.fasta_1.590
-DB=Issued Patents NA -QWTF=fastap -SUFFIX=rni -MINWATCH=0.1 -LOEPCI=0
-LOEPCI=0 -UNITS=bits -START=1 -END=1 -MATRIX=Dblom62 -TRANS=human40.cdi
-LIST=150 -DOALIGN=200 -THR SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=50
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10657740 @cgn 1 1 219 @runat_27052005_165254_3343 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:
5: /cgn2_6/ptodata/1/ina/6C.COMB.seq:
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	916	100.0	1114	4	US-09-949-016-5220
2	489	53.4	691	4	US-09-949-016-2545
3	489	53.4	856	4	US-09-513-999C-3891
4	489	53.4	893	4	US-09-513-999C-14942
5	489	53.4	911	4	US-09-513-999C-14938
6	489	53.4	927	4	US-09-513-999C-14936
7	489	53.4	942	4	US-09-513-999C-3890
8	457.5	49.9	913	4	US-09-513-999C-14937
9	429	46.8	7775	4	US-09-949-016-16962
10	406.5	44.4	826	4	US-09-621-976-3172
11	330	36.0	695	4	US-09-513-999C-3892
12	289.5	31.6	789	4	US-09-023-655-1300

13	289.5	31.6	845	4	US-09-949-016-4992	Sequence 4992, Ap
14	289.5	31.6	1380	3	US-09-513-783A-169	Sequence 169, App
15	283	30.9	1379	3	US-09-553-498-5	Sequence 5, Appli
16	283	30.9	1379	3	US-09-618-869-5	Sequence 5, Appli
17	230	25.1	601	4	US-09-949-016-182120	Sequence 182120,
18	230	25.1	7125	4	US-09-949-016-14287	Sequence 14287, A
19	182.5	19.9	393	4	US-08-513-999C-10503	Sequence 10503, A
20	171.5	18.7	1627	2	US-08-900-407-2	Sequence 2, Appli
21	163	17.8	509	4	US-09-513-999C-1048	Sequence 1048, Ap
22	163	17.8	546	4	US-09-513-999C-1051	Sequence 1051, Ap
23	163	17.8	564	4	US-09-513-999C-1047	Sequence 1047, Ap
24	163	17.8	580	4	US-09-513-999C-1049	Sequence 1049, Ap
25	163	17.8	595	4	US-08-513-999C-1046	Sequence 1046, Ap
26	158	17.2	601	4	US-09-949-016-91317	Sequence 91317, A
27	158	17.2	763	4	US-09-949-016-2623	Sequence 2623, Ap
28	158	17.2	4763	4	US-09-949-016-14365	Sequence 14365, A
29	149.5	16.3	333	4	US-09-513-999C-10502	Sequence 10502, A
30	147	16.0	5688	4	US-09-949-016-16734	Sequence 16734, A
31	142	15.5	601	4	US-09-949-016-89647	Sequence 89647, A
32	140.5	15.3	599	4	US-09-270-767-563	Sequence 563, App
33	140.5	15.3	599	4	US-09-270-767-15845	Sequence 15845, A
34	131.5	14.4	566	4	US-09-513-999C-1052	Sequence 1052, Ap
35	125.5	13.7	495	4	US-09-669-751-69	Sequence 69, Appli
36	118	12.9	601	4	US-09-949-016-89648	Sequence 89648, A
37	115	12.6	12214	4	US-09-949-016-17284	Sequence 17284, A
38	110.5	12.1	719	4	US-09-620-312D-669	Sequence 669, App
39	110	12.0	9615	4	US-09-949-016-17282	Sequence 17282, A
40	107	11.7	7125	4	US-09-949-016-14287	Sequence 14287, A
41	97	10.5	601	4	US-09-949-016-175645	Sequence 175645, A
42	96	10.5	420	4	US-09-107-532A-670	Sequence 670, App
43	93.5	10.2	652	1	US-07-748-761-2	Sequence 2, Appli
44	92	10.0	465	4	US-09-252-991A-12236	Sequence 12236, A
45	92	10.0	522	4	US-09-252-991A-12373	Sequence 12373, A
46	92	10.0	576	4	US-09-489-039A-3463	Sequence 3463, Ap
47	88	9.6	168971	4	US-09-949-016-13807	Sequence 13807, A
48	86	9.4	489	4	US-09-902-540-6436	Sequence 6436, Ap
49	86	9.4	3626	4	US-09-902-540-444	Sequence 444, App
50	84	9.2	49487	4	US-09-949-016-11770	Sequence 11770, A
51	84	9.2	57751	4	US-09-949-016-13631	Sequence 13631, A
52	82	9.0	20757	4	US-09-902-540-1189	Sequence 1189, Ap
53	79.5	8.7	1350	4	US-09-583-110-2262	Sequence 2262, Ap
54	79.5	8.7	1353	4	US-09-107-433-2322	Sequence 2322, Ap
55	79.5	8.7	9370	1	US-08-320-559-27	Sequence 27, Appl
56	79.5	8.7	9370	5	US-08-545-860D-27	Sequence 27, Appl
57	79.5	8.7	9370	5	PCT-US94-04496-27	Sequence 27, Appl
58	79.5	8.7	9391	1	US-08-320-559-25	Sequence 25, Appl
59	79.5	8.7	9391	5	US-08-545-860D-25	Sequence 25, Appl
60	79.5	8.7	9391	5	PCT-US94-04496-25	Sequence 25, Appl
61	77	8.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
62	77	8.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
63	76.5	8.4	510	4	US-09-489-039A-3514	Sequence 3514, Ap
64	76.5	8.4	1008	4	US-09-969-532-19	Sequence 19, Appl
65	76.5	8.4	1041	4	US-09-969-532-17	Sequence 17, Appl
66	76.5	8.4	1203	4	US-09-602-787A-149	Sequence 149, App
67	76.5	8.4	1701	4	US-09-969-532-3	Sequence 3, Appli
68	76.5	8.4	1734	4	US-09-969-532-1	Sequence 1, Appli
69	76.5	8.4	2010	4	US-09-969-532-27	Sequence 27, Appl
70	76.5	8.4	2043	4	US-09-969-532-25	Sequence 25, Appl
71	76.5	8.4	2055	4	US-10-029-180-81	Sequence 81, Appl
72	76.5	8.4	2703	4	US-09-969-532-11	Sequence 11, Appl
73	76.5	8.4	2736	4	US-09-969-532-9	Sequence 9, Appli
74	76.5	8.4	3411	1	US-09-969-532-33	Sequence 33, Appl
75	76	8.3	9432	1	US-08-277-231A-1	Sequence 1, Appli
76	76	8.3	9432	2	US-08-473-750-4	Sequence 4, Appli
77	76	8.3	9432	2	US-08-477-326-4	Sequence 4, Appli
78	75.5	8.2	966	4	US-09-969-532-23	Sequence 23, Appl
79	75.5	8.2	999	4	US-09-969-532-21	Sequence 21, Appl
80	75.5	8.2	1659	4	US-09-969-532-7	Sequence 7, Appli
81	75.5	8.2	1692	4	US-09-969-532-5	Sequence 5, Appli
82	75.5	8.2	1968	4	US-09-969-532-31	Sequence 31, Appl
83	75.5	8.2	2001	4	US-09-969-532-29	Sequence 29, Appl
84	75.5	8.2	2661	4	US-09-969-532-15	Sequence 15, Appl
85	75.5	8.2	2694	4	US-09-969-532-13	Sequence 13, Appl

```
86 75 8.2 360 4 US-09-543-681A-3563 Sequence 3563, Ap
c 87 74.5 8.1 34030 4 US-09-949-016-12248 Sequence 12248, A
c 88 74.5 8.1 34031 4 US-09-949-016-13540 Sequence 13540, A
89 73.5 8.0 712 4 US-09-833-909A-1 Sequence 1, Appli
90 73 8.0 633 4 US-09-252-991A-7328 Sequence 7328, Ap
91 73 8.0 771 4 US-09-252-991A-7403 Sequence 7403, Ap
92 73 8.0 1254 4 US-09-252-991A-7470 Sequence 7470, Ap
c 93 73 8.0 2103 4 US-09-252-991A-7175 Sequence 7175, Ap
94 73 8.0 3402 1 US-08-480-917-1 Sequence 1, Appli
95 73 8.0 3402 3 US-09-138-736-1 Sequence 1, Appli
96 73 8.0 3402 3 US-08-988-242-1 Sequence 1, Appli
c 97 73 8.0 62354 4 US-09-949-016-16188 Sequence 16188, A
98 72.5 7.9 978 4 US-09-248-796A-826 Sequence 826, App
99 72.5 7.9 3000 2 US-08-896-344A-1 Sequence 1, Appli
100 72.5 7.9 3000 3 US-09-360-682A-1 Sequence 1, Appli
101 72.5 7.9 50341 1 US-08-247-901C-1 Sequence 1, Appli
102 72.5 7.9 50341 2 US-09-075-904-1 Sequence 1, Appli
103 72.5 7.9 52297 3 US-09-426-436-1 Sequence 1, Appli
c 104 72.5 7.9 52297 3 US-08-705-557-1 Sequence 1, Appli
c 105 72.5 7.9 139936 4 US-09-949-016-11782 Sequence 11782, A
c 106 72.5 7.9 139952 4 US-09-949-016-13280 Sequence 13280, A
107 72 7.9 2041 4 US-09-465-559-13 Sequence 13, Appli
c 108 72 7.9 62908 4 US-09-949-016-17554 Sequence 17554, A
c 109 71.5 7.8 1354 3 US-09-252-991A-12170 Sequence 12170, A
c 110 71.5 7.8 2881 4 US-08-456-200B-4 Sequence 4, Appli
c 111 71.5 7.8 3563 4 US-09-949-016-5391 Sequence 5391, Ap
c 112 71.5 7.8 3563 3 US-09-041-886-20 Sequence 20, Appli
c 113 71.5 7.8 3596 2 US-08-779-801-5 Sequence 5, Appli
c 114 71.5 7.8 3596 3 US-09-298-441-5 Sequence 3, Appli
c 115 71.5 7.8 3632 2 US-08-779-801-3 Sequence 4, Appli
c 116 71.5 7.8 3632 3 US-09-298-441-4 Sequence 4, Appli
c 117 71.5 7.8 3632 3 US-09-298-441-3 Sequence 3, Appli
c 118 71.5 7.8 7791 2 US-08-149-097D-23 Sequence 23, Appli
c 119 71.5 7.8 7791 3 US-08-949-386-23 Sequence 23, Appli
c 120 71.5 7.8 7791 3 US-08-450-562-23 Sequence 23, Appli
c 121 71.5 7.8 7791 3 US-08-984-709A-23 Sequence 23, Appli
c 122 71.5 7.8 7791 3 US-08-450-272-23 Sequence 23, Appli
c 123 71.5 7.8 7791 3 US-08-450-272-23 Sequence 23, Appli
c 124 71.5 7.8 7808 2 US-08-149-097D-22 Sequence 22, Appli
c 125 71.5 7.8 7808 3 US-08-949-386-22 Sequence 22, Appli
c 126 71.5 7.8 7808 3 US-08-450-562-22 Sequence 22, Appli
c 127 71.5 7.8 7808 3 US-08-984-709A-22 Sequence 22, Appli
c 128 71.5 7.8 7808 3 US-08-450-272-22 Sequence 22, Appli
c 129 71.5 7.8 7808 4 US-08-450-272-22 Sequence 22, Appli
c 130 71.5 7.8 7815 4 US-08-450-272-22 Sequence 22, Appli
c 131 71.5 7.8 7815 4 US-09-949-016-3629 Sequence 3629, Ap
c 132 71.5 7.8 7815 4 US-09-949-016-3630 Sequence 3630, Ap
c 133 71.5 7.8 70559 4 US-09-409-800B-1 Sequence 1, Appli
c 134 71.5 7.8 1230230 4 US-09-438-185A-1 Sequence 1, Appli
135 71 7.8 627 4 US-09-601-198-113 Sequence 113, App
c 136 71 7.8 1230 4 US-09-489-039A-6381 Sequence 6381, Ap
c 137 71 7.8 1278 4 US-09-902-540-8457 Sequence 8457, Ap
c 138 71 7.8 3762 4 US-09-435-376-1 Sequence 1, Appli
c 139 71 7.8 8590 4 US-09-902-540-889 Sequence 889, App
140 71 7.8 34794 4 US-09-713-678-39 Sequence 39, Appli
c 141 71 7.8 4403765 3 US-09-103-840A-2 Sequence 2, Appli
c 142 71 7.8 4411529 3 US-09-103-840A-1 Sequence 1, Appli
c 143 70.5 7.7 585 4 US-09-248-796A-2425 Sequence 2425, Ap
144 70.5 7.7 1605 3 US-09-124-541-3 Sequence 3, Appli
c 145 70.5 7.7 1605 4 US-09-663-326-3 Sequence 3, Appli
c 146 70.5 7.7 5816 3 US-08-857-076-11 Sequence 11, Appli
c 147 70.5 7.7 1230025 4 US-09-198-452A-1 Sequence 1, Appli
c 148 70 7.6 601 4 US-09-949-016-142871 Sequence 142871,
c 149 70 7.6 2247 4 US-09-252-991A-5125 Sequence 5125, Ap
150 70 7.6 2952 4 US-09-252-991A-5096 Sequence 5096, Ap
```

ALIGNMENTS

RESULT 1

```
US-09-949-016-5220
; Sequence 5220, Application US/09949016
; Patent No. 6812339
```

RESULT 2

```
US-09-949-016-2545
; Sequence 2545, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5220
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5220
```

```
Alignment Scores:
Pred. No.: 1,7e-120 Length: 1114
Score: 916.00 Matches: 173
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
```

US-10-657-740-1 (1-173) x US-09-949-016-5220 (1-1114)

```
Qy 1 MetAspValThrIleGlnHisProTrrPheIysArgThrLeuGlyProPheTyrProSer 20
Db 70 ATGACGTGACCATCCAGCACCCCTGTTCAAGCGCACCTTGGGGCCCTTCTACCCGACG 129
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
Db 130 CGCTGTTCGACCACTTTTTCGGCGAGGGCTTTTTCAGTATGACCTGCTGCCCTTCTG 189
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 190 TCGTCCACCATCAGCCCTACTACCGCAGTCCCTCTTCGCGACCGTGTGGACTCCGGC 249
Qy 61 IleSerGluValArgSerAspArgAspIysPheValIlePheLeuAspValIysHisPhe 80
Db 250 ATCTCTGAGGTTCGATCCGACCGGGAACAAGTTCGTCTCATCTTCCTCGATGTGAAGCATT 309
Qy 81 SerProGluAspLeuThrValIysValIcInAspAspPheValGluIleHisGlyLysHis 100
Db 310 TCCCGGAGGACCTCCCGTGAAGTGCAGGACGACTTTGTGGAGATCCACGGAAGCAC 369
Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu 120
Db 370 AACGAGCGCCAGGACGACCAACCGGTACATTTCCGCTGAGTTCCACCGCGCTACCGCTG 429
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 430 CCGTCCAACTGGACCATCGGCCCTCTTGTCTCTCTGTCGCGATGGCATGCTGACC 489
Qy 141 PheCysGlyProIysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 490 TTCTGTGGCCCAAGATCCAGACTGGCTGGATGCCACCCACCGCGGAGGACCATCCCC 549
Qy 161 ValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 550 GTGTCCGGGAGGAGAACCCACCTCGGCTCCCTCGTCC 588
```

;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2545
;; LENGTH: 691
;; TYPE: DNA
;; ORGANISM: Human
US-09-949-016-2545

Alignment Scores:
Pred. No.: 7,89e-60 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 4 Gaps: 6

US-10-657-740-1 (1-173) x US-09-949-016-2545 (1-691)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 26 ATGGACATCGCCATCCACCCCTGGATCGCGCCCTCTTCTTCCATCCACCC 85
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 86 AGCCGCCCTTTGACAGTTCTTCGAGAGACCTTGTGGAGTCTGATCTTTTCCCG--- 142
QY 40 LeuSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 143 ACGTCTACTTCCCTGAGTCCCTTCTACTCTCGGCCACCTCTCTCTCGGGCACCAGC 202
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 203 TGGTTTGACATCGGACTCTCAGAGATGCGCTGGAGAGAGACAGTTCTCTGTCAACCTG 262
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
DB 263 GATGTGAAGCACTTCTCCCGAGGAACTCAAAGTTAAGGTGTGGAGATGTGATTGAG 322
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 323 GTGCATGAAACATGAAGAGCGCCAGCATGACATGATGTTTCATCTCCAGGAGTTCAC 382
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 383 AGGAATACCGGATCCCGATGATGATAGACCTCTCACCATTACTTCTCCCTGTCTCATCT 442
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 443 GATGGGGCTCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCT 490
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 491 GAGCGCACCATTTCCCATCACCCTGAGAGAGACCTGTGTGTCACCGCAGCCCC 544

RESULT 3

US-09-513-999C-3891
; Sequence 3891, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.

;; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
;; Patent No. 6783961
;; FILE REFERENCE: 59.US2.REG
;; CURRENT APPLICATION NUMBER: US/09/513,999C
;; CURRENT FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/122,487
;; PRIOR FILING DATE: 1999-02-26
;; NUMBER OF SEQ ID NOS: 36681
;; SOFTWARE: Patent.pm
;; SEQ ID NO 3891
;; LENGTH: 856
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 4..753
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 51
;; OTHER INFORMATION: s=g or c
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 117
;; OTHER INFORMATION: s=g or c
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 118
;; OTHER INFORMATION: k=g or t
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 159
;; OTHER INFORMATION: r=a or g
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 160
;; OTHER INFORMATION: s=g or c
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 182
;; OTHER INFORMATION: r=a or g
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: 183
;; OTHER INFORMATION: r=a or g
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: 16
;; OTHER INFORMATION: Xaa=Asp or Glu
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: 38
;; OTHER INFORMATION: Xaa=His or Gln
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: 39
;; OTHER INFORMATION: Xaa=Ala or Ser
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: 53
;; OTHER INFORMATION: Xaa=Glu or Gln
;; FEATURE:
;; NAME/KEY: UNSURE
;; LOCATION: 60
;; OTHER INFORMATION: Xaa=Glu or Gly
US-09-513-999C-3891

Alignment Scores:
Pred. No.: 1.12e-59 Length: 856
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 4 Gaps: 6

```
US-10-657-740-1 (1-173) x US-09-513-999C-3891 (1-856)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 229 ATGGACATGCCATCCACACCCCTGGATCCGCGCCCTTCTTTCTTCCACTCCCCC 288
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 289 AGCCGGCTCTTTGACAGTCTTCGGAGAGACCTGTTGGAGTCTGATCTTTTCCCG--- 345
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 346 AGCTCTACTTCCCTGAGTCCCTTACTCTTCGCCACCCCTCTCTTCCTGGGGACCCAGC 405
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 406 TGGTTTGACACTGGACTCTCAGAGATCGCCTGGAGAGACAGGTTCTCTGTCAACCTG 465
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheValGlu 95
Db 466 GATGTGAAGACATTCCTCCAGAGAACTCAAAGTTAAGGTGTTGGAGATGTGATTGAG 525
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 526 GTGCATGGAACATGAAGAGCGCCAGGATGAACATGTTTCATCTCCAGGGAGTTCCAC 585
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 586 AGGAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTACTTTCATCCCTGTCTCT 645
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 646 GATGGGGTCTCTACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCT 693
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 694 GAGCGCACCATTCCTCCATCACCCTGAAGAGAGCCCTGTCTCACCAGCCGCC 747

RESULT 4
US-09-513-999C-14942
; Sequence 14942, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14942
; LENGTH: 893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14942

Alignment Scores:
Pred. No.: 1.19e-59 Length: 893
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 4 Gaps: 6

US-10-657-740-1 (1-173) x US-09-513-999C-14942 (1-893)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 284 ATGGACATGCCATCCACACCCCTGGATCCGCGCCCTTCTTTCTTCCACTCCCCC 343
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 284 ATGGACATGCCATCCACACCCCTGGATCCGCGCCCTTCTTTCTTCCACTCCCCC 343

US-10-657-740-1 (1-173) x US-09-513-999C-14938 (1-911)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 284 ATGGACATGCCATCCACACCCCTGGATCCGCGCCCTTCTTTCTTCCACTCCCCC 343
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 284 ATGGACATGCCATCCACACCCCTGGATCCGCGCCCTTCTTTCTTCCACTCCCCC 343
```

Db 344 AGCGGCTCTTGGACAGTCTTGGAGAGCAGCTTGGAGTCTGATCTTTCCCG--- 400
Qy 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55
Db 401 AGCTCTACTTCCCTGAGTCCCTTACTCTCGGCCACCCCTCTTCCCTCGGGGACCCAGC 460
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspIlePheValIlePheLeu 75
Db 461 TGGTTTGACACTGGAGTCTCAGAGATGCGCTCGGAGAGGACAGGTTCTCTGTCAACCTG 520
Qy 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGlu 95
Db 521 GATGTGAACACTTCTCCAGAGGAACTCAAGTTTAAGTTGTTGGAGATGTGATTGAG 580
Qy 96 IleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115
Db 581 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCCAC 640
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 641 AGGAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTACTTCCCTGTCTATCT 700
Qy 136 AspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 701 GATGGGTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCT 748
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 749 GAGCGCACCATTCCCATCACCCTGAGAGAGACCTGTGTCCAGCCGACCCCC 802

RESULT 6

US-09-513-999C-14936
; Sequence 14936, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14936
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-14936

Alignment Scores:
Pred. No.: 1,278-59 Length: 927
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 4 Gaps: 6

US-10-657-740-1 (1-173) x US-09-513-999C-14936 (1-927)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 300 ATGGACATCGGCATCCACACCCCTGGATCGCCGCCCTTCTTCTTCCACTCCCC 359
Qy 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuPheProPhe 39
Db 360 AGCCGCTCTTGGACAGTCTTCCGAGAGACCTGTGGAGTCTGATCTTTTCCG--- 416
Qy 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55

Db 417 AGCTCTACTTCCCTGAGTCCCTTACCTTGGCCACCCCTCTCTCTCGGGGACCCAGC 476
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspIlePheValIlePheLeu 75
Db 477 TGGTTTGACACTGGAGTCTCAGAGATGCGCTCGGAGAGGACAGGTTCTCTGTCAACCTG 536
Qy 76 AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspPheValGlu 95
Db 537 GATGTGAAGACACTTCTCCGAGAGAACTCAAGTTTAAGTTGTTGGAGATGTGATTGAG 596
Qy 96 IleHisGlyIleHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHis 115
Db 597 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCCAC 656
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 657 AGGAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTACTTCCCTGTCTATCT 716
Qy 136 AspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 717 GATGGGTCTCTCACTGTGAATGACCAAGAAACAG-----GTCTCTGGCCCT 764
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db 765 GAGCGCACCATTCCCATCACCCTGAGAGAGACCTGTGTCCAGCCGACCCCC 818

RESULT 7

US-09-513-999C-3890
; Sequence 3890, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 3890
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 315..839
US-09-513-999C-3890

Alignment Scores:
Pred. No.: 1,38e-59 Length: 942
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 4 Gaps: 6

US-10-657-740-1 (1-173) x US-09-513-999C-3890 (1-942)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 315 ATGGACATCGGCATCCACACCCCTGGATCGCCGCCCTTCTTCTTCCACTCCCC 374
Qy 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuPheProPhe 39
Db 375 AGCCGCTCTTGGACAGTCTTCCGAGAGACCTGTGGAGTCTGATCTTTTCCG--- 431
Qy 40 LeuSerSerThrIleSerProTyrArgGln-----SerLeuPheArg-----Thr 55

```
Db 432 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCTCTCTCTCTGCGGGCACCACG 491
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
492 TGGTTGACACTGGACTCTCAGATGCGCCTGGAGAGGACAGGTTCTCTGTCAACCTG 551
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
552 GATGTGAACCACTTCTCCCGCAGAGAACTCAAGTTAAGTCTTGGAGATGTGATTGAG 611
Qy 96 IleHisGlyLysHisAsnGlnArgGlnAspAspHisGlyTyIleSerArgGluPheHis 115
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
612 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCCAGGGAGTTCCAC 671
Qy 116 ArgArgTyIleSerProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
672 AGGAATACCGATCCGATCCAGCTGATGTAGACCTCTCACCATTTACTTCATCTGTCTCT 731
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
732 GATGGGTCTCTACTGTGAATGACCAAGGAACAG-----GTCTCTGGCCCT 779
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
780 GAGCGCACCATTTCCCATCCCGTGAAGAGAGCCTGTGTCTCAGCGCAGCCCC 833

RESULT 8
US-09-513-999C-14937
; Sequence 14937, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Malne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14937
; LENGTH: 913
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 203
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 204
; OTHER INFORMATION: n=a, g, c or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 330
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 372
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 390
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 394
; OTHER INFORMATION: s=g or c
US-09-513-999C-14937
```

```
Alignment Scores:
Pred. No.: 3,91e-55 Length: 913
Score: 457.50 Matches: 94
Percent Similarity: 71.91% Conservative: 34
Best Local Similarity: 52.81% Mismatches: 33
Query Match: 49.95% Indels: 17
DB: 4 Gaps: 7
US-10-657-740-1 (1-173) x US-09-513-999C-14937 (1-913)
Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
301 ATGCATCGCAGTCCACCCACCTTC-----TTKCTTTTCCACTCCGCC 345
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyIleAspLeuProPhe 39
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
346 AGCCGCTCTTTTGACCACTTCTTCGCGAGACCTGTTGGAGTCKGATSTTTTCCCG--- 402
Qy 40 LeuSerSerThrIleSerProTyIleArgGln-----SerLeuPheArg-----Thr 55
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
403 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCTCTCTTCTGGGGCACCACG 462
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
463 TGGTTGACACTGGACTCTCAGATGCGCCTGGAGAGGACAGGTTCTCTGTCAACCTG 522
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
523 GATGTGAAGCACTTCTCCCGCAGAGAACTCAAGTTAAGGTGTTGGGAGATGTGATTGAG 582
Qy 96 IleHisGlyLysHisAsnGlnArgGlnAspHisGlyTyIleSerArgGluPheHis 115
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
583 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGGTTTCATCTCAGGGAGTTCCAC 642
Qy 116 ArgArgTyIleSerProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
643 AGGAATACCGATCCCGATCGATGTAGACCTCTCACCATTTACTTCATCTCTGTCTCT 702
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
703 GATGGGTCTCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCT 750
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
751 GAGCGCACCATTTCCCATCCCGTGAAGAGAGCCTGTGTCTCAGCGCAGCCCC 804

RESULT 9
US-09-949-016-16962
; Sequence 16962, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16962
; LENGTH: 7775
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16962
Alignment Scores:
Pred. No.: 1.48e-49 Length: 7775
```

Score: 429.00 Matches: 172
Percent Similarity: 16.23% Conservative: 0
Best Local Similarity: 16.23% Mismatches: 1
Query Match: 46.83% Indels: 888
DB: 4 Gaps: 2

US-10-657-740-1 (1-173) x US-09-949-016-16962 (1-7775)

```
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyrProSer 20
DB 2070 ATGGAGCTGACCATCCAGCACCCCTGGTTCAAGGCGCACCTGGGGCCCTTCTACCCAGC 2129
QY 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
DB 2130 CGGCTGTTCGACAGTTTTTTCGGCGAGGGCCCTTTTGTAGTATGACCTGTGCTGCTTCTCG 2189
QY 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
DB 2190 TCGTCCACCATCAGCCCCCTACTACGCCAGTCCCTTTCGCCACCGCTGTGACTCCGGC 2249
QY 61 IleSerGluVal----- 64
DB 2250 ATCTCTGAGGT-AAGACGTGGCCCGTGGTGGCTCTCTCTCGCTGCTCAGAGGGTGT 2308
QY 64 ----- 64
DB 2309 GGCCTCGTGGTGGAGAGCATGGACTCTGTCTTGTCTTCTCGTCCGTCAGGCGAGTGGCTCGT 2368
QY 64 ----- 64
DB 2369 CCCACTTCATCCCTTTCAGAGGCTGGGCGAGAGCGTGTGTCCCACTGCAGCCAGCTGG 2428
QY 64 ----- 64
DB 2429 CAGAGCTTCCCTGGCACTGGGGAGAGGGTGGCAAGGGAGCAGCTGTAATCCACTTTG 2488
QY 64 ----- 64
DB 2489 CTTTCTCCTCATCAGTCTATGACCCATGTGTGTCTTTGTAGGACCAAGCCACATCTGGAA 2548
QY 64 ----- 64
DB 2549 ACCCAAGAGCAGCCCATCCAGGCATGCGTGTGGCGAATGCCAGTCCCGGGTCTCT 2608
QY 64 ----- 64
DB 2609 GGTCTCTGAGTCCCGGAGACCTGGGAGCAGGTGGGGGTATAGTCTGAAAGCCAGAGA 2668
QY 64 ----- 64
DB 2669 GCAGGCGGTTCTAGCACCTCTCCAAATGAGCTCGGCTGCCCCACGGCTAGCAAAGCTCT 2728
QY 64 ----- 64
DB 2729 TGGCAAGTTTACTTAGGTGCCCTGCCAAGGCTAAAGGACAGGCAATGGACGCCCGCCCC 2788
QY 64 ----- 64
DB 2789 CCCCACCCNAACACAGGCCCTCTCTGTAGCCACGGGTGAGCGGTGTCAGGTTCTGCTGTT 2848
QY 64 ----- 64
DB 2849 CTGGAGGGCTGAGTCCCAACCCAGCACCTCATAAACAGGGTCTCTCCAGGGCTGTGCA 2908
QY 64 ----- 64
DB 2909 GTAGGCATCAATGCCAGGGTGCAAAATGCTCTCAGGAGCCAAGGCTGAGCCAGGGAGTG 2968
QY 64 ----- 64
DB 2969 AGAAGAGCATGTGGAGTGGCTTTTGGAGAGGAGCTCGGAGGCTGTGACGAGCTCC 3028
QY 64 ----- 64
```

```
DB 3029 GGCCGCTTCTATAGACAGCATGACACCAAGGGCAGTGAACCTCATTTCCACAGGCTGAGTCC 3088
QY 64 ----- 64
DB 3089 AGCCAGCAGCCCAAGCATCACCAGCCAGACGATTGACCTTAAGGGACCAACCAACCCGTA 3148
QY 64 ----- 64
DB 3149 ACGACCCCTCTTACCATAACAGTAGCCAGCCAGCCAGCCATAACCAAGCAACTTATCTATAA 3208
QY 64 ----- 64
DB 3209 CCAGCCACTGTACCATAGCCAAACACAGCCGGCCCAACCATAGCATTTACGCCCTCAG 3268
QY 64 ----- 64
DB 3269 CTGGCCCTGAGGGTTTGGAGACAGGTCGAGGGTCATGCTGTCTGTCCAGGAGACAGTCA 3328
QY 64 ----- 64
DB 3329 CAGGCCCCGAAAGCTCTGCCCACTTGTGTGTGGAGAGAGGCGCGCAGGTGACCGA 3388
QY 64 ----- 64
DB 3389 AGCATCTCTGTTCGATAACCGGGACCCGCCCTGTCTCTGCCAACCCACAGGGGACGGC 3448
QY 65 -----ArgSerAspArgAspLys 70
DB 3449 ACCCTCTGGGCGCTCCATGGCACGTTTGGATTTCAGTTTCGATCCGACCGGACAG 3508
QY 71 PheValIlePheLeuAspValIlyHisPheSerProGluAspLeuThrValIlyValGln 90
DB 3509 TTGTCATCTCTCTCGATGTGAAGCACTTCTCCCGGAGGACCTCACCGTGAAGGTGCAG 3568
QY 91 AspAspPheValGluIleHisGlyLysHisArgGluArg----- 103
DB 3569 GACGACTTTGTGGAGATCCACGAAAGCACACACGAGCGCCAGGTGAGCCACAGCACTGAG 3628
QY 103 ----- 103
DB 3629 AGTGGGAGAGGGGGCGAGTTGGGCGGAGGACAAGGGGTTCACGGCGGGCACGACCGG 3688
QY 103 ----- 103
DB 3689 GCCTGCACACCTGCACCATGCTTCAACCTGGGAGAGGAGCGCTCTCCAGGGGACCCCG 3748
QY 103 ----- 103
DB 3749 AATCAGGCTTGGCTTTTCCCAAGGGAGGGGCGCATGCCCACTGAGCACAGCCAGCCCT 3808
QY 103 ----- 103
DB 3809 CCCGCTGACAGAGGTACCATTTCCCGAGCTAATGTGGCTCAGGGATCCAGGTTAGGGTCC 3868
QY 103 ----- 103
DB 3869 CTTCCCGGCTGCACCCAGCGCTGCCAGCTCCATCCCTGTCACTCTGATGCCAGGGTGG 3928
QY 103 ----- 103
DB 3929 TCTTAGAAGAAACCCAGGAAAGTGGAGTGGCCCGGTGGCGCCTCTTAGCCAGTGTATC 3988
QY 103 ----- 103
DB 3989 ATCTTCATGAACCTTACCTAGGAAAGCAGTCCCGACGGCATAGCTGATCCGCTTG 4048
QY 103 ----- 103
DB 4049 GAATGCTTTTACAGGCAATTGACACCTTTCGCTTCACAGCAGCACTTTTGGAAACAGTGTCTC 4108
QY 103 ----- 103
```

Db 4109 ATTATTCAGGGGACGGCTGGGGAAACAAGGGGGTCTCTCAGCGCTGCTGGGTCCCAAGCTA 4168
Qy 103 ----- 103
Db 4169 GTACGGGACAGTGGACGGAGCTTCTCCCAACAGTCACTCCATGATGCCCGCTCTTGCTC 4228
Qy 103 ----- 103
Db 4229 GGCTGGAGGCTCGGATCTCCGTGGTGTGAGGAGCGGGGCACTGAGGCCCTGGTGAC 4288
Qy 103 ----- 103
Db 4289 CTGCATCTCTGGCGGACCGGAGAGCTCATGAGCTGTACAGATGGACAGTGCCTCCG 4348
Qy 103 ----- 103
Db 4349 CGGGGCTGGAGACAGAGTGGGGCTGGAGGTGGAACCTTAGCCAAAGTCTTTGGTTTC 4408
Qy 103 ----- 103
Db 4409 TTTTGGCAGGGTCTCTTTCAATGGCTGGAGAGGTGGTCTGGGGGTGAACGCTGAC 4468
Qy 103 ----- 103
Db 4469 CTCCTCATGTGTCCTCCCTCGCTGGGCGCGGTAAAGCCCGACGTAGCCCCAGCC 4528
Qy 103 ----- 103
Db 4529 AGCTGGAACATGTTCTCTGAGCTCCCAAGCTCTTGGTCTTTGGACCCAGTGGAGGAG 4588
Qy 103 ----- 103
Db 4589 GTCAGCCAGGAGCTGAGTCTGCGGTTTAGGGCGTCCAGGGAGCTGGAAGCATGTGG 4648
Qy 103 ----- 103
Db 4649 TCGTCTGCCACATTAGGTAGGGCTGCAGAGACCTGGGCTAGAGCAGTCTCGGGGTCT 4708
Qy 103 ----- 103
Db 4709 GGAAGGGGAAGACTGGCTGAGGTGGGGGCTGTGGAATGATTCCTGCGATTTGGAG 4768
Qy 103 ----- 103
Db 4769 TGAAGCCATGAGCGGGAGAGACAACCCCGCGGGGAATAGCCCGCAAGTGGCCACG 4828
Qy 103 ----- 103
Db 4829 AGGCAGGCTGAGGTCCAGAGAAGCAGGGGCGATGAATCCATAAATCCAGGGGGCTGGC 4888
Qy 103 ----- 103
Db 4889 CATGGATGTCTGGCTGCACCCCGCCCTGTGAGAGCCCCCGCAGGCTGGCCCCCTTCT 4948
Qy 103 ----- 103
Db 4949 CGAGTCAGTGGGGCTGGGGCAGCTTCTCTGCGATGGGGGAGGAGCGCGCTGCACAGTG 5008
Qy 104 -----GlnAspHisGlyTyrIleSerArgGlu 113
Db 5009 GCGCCCTGACTGTGCGCCCCCACCCTCTCCAGAGCAGCACCGCTACATTTCCGCTGAG 5068
Qy 114 PheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeu 133
Db 5069 TTCCACGCGCCCTACCGCTCCCGTCCCAACGTGAGCCAGTCCGCGCCCTCTCTTGTCTCCCTG 5128
Qy 134 SerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThr 153
Db 5129 TCTGCGGATGCATGCTGACCTTCTGTGGCCCCCAAGATCCAGACTGGCGCTGGATGCCACC 5188
Qy 154 HisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
Db 5189 CACGCGAGCGAGCATCCCGTGTGCGGGGAGGAGAGCCACCTCGGCTCCCTCGTCC 5248

RESULT 10

US-09-621-976-3172
; Sequence 3172, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3172
; LENGTH: 826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 557..736
US-09-621-976-3172

Alignment Scores:
Pred. No.: 6.43e-48 Length: 826
Score: 406.50 Matches: 93
Percent Similarity: 68.72% Conservative: 30
Best Local Similarity: 51.96% Mismatches: 37
Query Match: 44.38% Indels: 19
DB: 4 Gaps: 5

US-10-657-740-1 (1-173) x US-09-621-976-3172 (1-826)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeu-GlyProPheTyr---Pr 19
Db 267 ATGGACATCGCATCCACACCCCTTGATCGGCAGSCCTTCTTCTCTTCCACTCCCC 326
Qy 19 oSerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAsp----- 35
Db 327 CAGCGCGCTCTTTGACAGTTCTTCGGAGAGCACCTGTGGAGTCTGATATTTCCCG 386
Qy 36 ----LeuLeuProPheLeuSerSerThrIleSer-----ProTyrTyrArgGlnSerLe 52
Db 387 ACGTCTACTTCCCTGAGTCTCTTCTACCTTCGGCCACCTCTTACCTCGCGGCACCCAG 446
Qy 52 uPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheVa 72
Db 447 CTGG-----TTTGACACTGGACTCTCAGAGATCGCTGGAGAGGACAGGTTCTC 497
Qy 72 lIlePheLeuAspValIleHisPheSerProGluAspLeuThrValIleValGlnAspAs 92
Db 498 TGTCACCTGGATGTGAAGCACCTTCTCCAGAGGAACCTCAAAGTTAAGGTGTGGGAGA 557
Qy 92 pPhe-ValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerA 112
Db 558 TGTGATTTAGGTGCATGGAACATGAGAGCGCCAGGATGAACATGTTTCTATCTCCA 617
Qy 112 rgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCys 132
Db 618 GGGAGTTCCACAGAGAAATACCGGATCCAGCTGATGTAGACCCCTCTCACCATTACTTCA 677
Qy 132 erLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspA 152
Db 678 CCCTGTCTATGTGTTGGGTCTCTCACTGTGAATGACCAAGGAAACAG-----G 725
Qy 152 laThr-HisAlaGluArgAlaIleProValSer-ArgGluGluLysPro 167
Db 726 TCTCTGGGCGCTGAGCGCACCATTCCTCCATCACCCCGTGAAGAGAGCCT 774

RESULT 11

US-09-513-999C-3892
; Sequence 3892, Application US/09513999C

Db 347 CTGGATGTCACCACTTCGCCCGGACGAGCTGACGGTCAAGACCAAGATGGCGTGTG 406
Qy 95 GluileHisGlyIysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPhe 114
Db 407 GAGATCATCCGCAACGACGAGGCGGACGAGCATGCTACATCTCCCGGTGCTTC 466
Qy 115 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
Db 467 ACGGGAAATACACGTCGCCCGGTGTGACCCCAAGTTCTCTCTCCTGTCC 526
Qy 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
Db 527 CCTGAGGACACATGACCGTGA-GGCCCC-----CAT 558
Qy 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
Db 559 GCCCAAGCTAGCCAGCGACTCAACGAGATCACCATCCCACT 600

RESULT 13

US-09-949-016-4992
; Sequence 4992, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4992
; LENGTH: 845
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4992

Alignment Scores:
Pred. No.: 3,46e-31 Length: 845
Score: 289.50 Matches: 75
Percent Similarity: 50.52% Conservative: 23
Best Local Similarity: 38.66% Mismatches: 43
Query Match: 31.60% Indels: 54
DB: 4 Gaps: 6

US-10-657-740-1 (1-173) x US-09-949-016-4992 (1-845)

Qy 10 PheLysArgThrLeuGlyPro-----PheTyrPro---SerArg 21
Db 129 TTCTCGCTCTGCGGGGCCCCAGCTGGGACCCCTTCGCGACTGTGATCCCGCATAGCGC 188
Qy 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu--- 40
Db 189 CTCTTCGACCAAGCGCTTCGGG-----CTGCCCGGCTGCGC 224
Qy 40 ----- 40
Db 225 GAGGAGTGGTCCAGTGGTTAGCGGCGACAGCTGGCCAGGCTACGTGCGGCCCTGCCCC 284
Qy 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
Db 285 CCCGCGGCATCGAGAGCCCGCAGTGGCGGCGCCGCTACAGCGCGCGCTAGCGCG 344
Qy 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 74
Db 345 CAA---CTCAGCAGCGGGGTCTCGGAGATCGGCACATCGGCACACCGTGGCGGTGTC 401

Qy 75 LeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspPheVal 94
Db 402 CTGGATGTCACCACTTCGCCCGGACGAGCTGACGGTCAAGACCAAGATGGCGTGTG 461
Qy 95 GluileHisGlyIysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPhe 114
Db 462 GAGATCATCCGCAACGACGAGGCGGACGAGCATGCTACATCTCCCGGTGCTTC 521
Qy 115 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
Db 522 ACGGGAAATACACGTCGCCCGGTGTGACCCCAAGTTCTCTCTCCTGTCC 581
Qy 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
Db 582 CCTGAGGACACATGACCGTGA-GGCCCC-----CAT 613
Qy 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
Db 614 GCCCAAGCTAGCCAGCGACTCAACGAGATCACCATCCCACT 655

RESULT 14

US-09-513-783A-169
; Sequence 169, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 169
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP-HSP27
; NAME/KEY: CDS
; LOCATION: (1)..(1380)
US-09-513-783A-169

Alignment Scores:
Pred. No.: 7,65e-31 Length: 1380
Score: 289.50 Matches: 75
Percent Similarity: 50.52% Conservative: 23
Best Local Similarity: 38.66% Mismatches: 43
Query Match: 31.60% Indels: 54
DB: 3 Gaps: 6

US-10-657-740-1 (1-173) x US-09-513-783A-169 (1-1380)

Qy 10 PheLysArgThrLeuGlyPro-----PheTyrPro---SerArg 21
Db 784 TTCTCGCTCTGCGGGGCCCCAGCTGGGACCCCTTCGCGACTGTGATCCCGCATAGCGC 843
Qy 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu--- 40
Db 844 CTCTTCGACCAAGCGCTTCGGG-----CTGCCCGGCTGCGC 879
Qy 40 ----- 40
Db 880 GAGGAGTGGTCCAGTGGTTAGCGGCGACAGCTGGCCAGGCTACGTGCGGCCCTGCCCC 939
Qy 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
Db 940 CCCGCGGCATCGAGAGCCCGCAGTGGCGGCGCCGCTACAGCGCGCGCTCAGCGCG 999
Qy 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 74
Db 1000 CAA---CTCAGCAGCGGGGTCTCGGAGATCGGCACATCGGCACACCGTGGCGGTGTC 1056


```
Db 671 ACCCTGGCGCCAGCAGCTTACGCGAGCGCTCAACCGACAG---CTCAGCAGCGGGGTC 727
Qy 62 SerGluValArgSerAspArgAspLysPheValIlePheLeuAspValIleHisPheSer 81
Db 728 TCGGAGATCCGACAGACAGCGCTGATCGCTGGCGCGTGTCCCTGGAGCTCAACCACTTCGCT 787
Qy 82 ProGluAspLeuThrValIleValGlnAspPheValGluIleHisGlyIleHisAsn 101
Db 788 CCGAGGAGCTCACAGTGAAGAACGAAGGCGGTGGAGATCACTGGCAAGCAGCAA 847
Qy 102 GluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuPro 121
Db 848 GAAAGCGAGGACGACATCGCTACATCTCGGTCTTCACCGGAAATACACGCTCCCT 907
Qy 122 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 141
Db 908 CCAGGTGTGGACCCACCGCTAGTGTCTTCCCTATCCCTGAGGGGACACCTTACCGTG 967
Qy 142 CysGly-----ProIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIle 159
Db 968 GAGGCTCCGTGTCGCAAGCAGTACG-----CAGTCAGCGGAGATCACCAATT 1015
Qy 160 ProValSerArgGluGluLys 166
Db 1016 CCGGTACTTTCAGGCGCCGC 1036

RESULT 17
US-09-949-016-182120
; Sequence 182120, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 182120
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-182120

Alignment Scores:
Pred. No.: 6,31e-23 Length: 601
Score: 230.00 Matches: 45
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.11% Indels: 0
DB: 4 Gaps: 0

US-10-657-740-1 (1-173) x US-09-949-016-182120 (1-601)

Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 2 CGGCTGTTTCGACCACTTTTCGGCGAGGCGCTTTTGAGTATGACCTGCTGCCCTTCCTG 61
Qy 41 SerSerThrIleSerProTyrArgGlnSerLeuPheArgThrValLeuAspSerCly 60
Db 62 TCGTCCACATCAGGCCCCCTACTACCGCCAGTCCCTCTTCCGACCGTGTGACTCCGGC 121
Qy 61 IleSerGluValArg 65
Db 122 ATCTCTGAGGTAGA 136
```

RESULT 18

```
US-09-949-016-14287
; Sequence 14287, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14287
; LENGTH: 7125
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(7125)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14287

Alignment Scores:
Pred. No.: 3,45e-15 Length: 7125
Score: 188.00 Matches: 37
Percent Similarity: 75.36% Conservative: 15
Best Local Similarity: 53.62% Mismatches: 11
Query Match: 20.52% Indels: 6
DB: 4 Gaps: 2

US-10-657-740-1 (1-173) x US-09-949-016-14287 (1-7125)

Qy 105 AspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeuProSerAsnVal 124
Db 4783 GATGAACATGTTTCATCTCCAGGGAGTTCACAGGAAATACCGGATCCAGCTGATGTA 4842
Qy 125 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 144
Db 4843 GACCTCTCACATTACTTCTCATCTGTCATCTGATGGGTCTCTCACTGTGAATGGACCA 4902
Qy 145 LysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGlu 164
Db 4903 AGGAACAG-----GTCTGGCCCTGAGCGCCACCATTCATCCACCCGTGAA 4950
Qy 165 GluLysPro-----ThrSerAlaPro 171
Db 4951 GAGAAGCTGCTGTCTCACCGCAGCCGCC 4977

RESULT 19
US-09-513-999C-10503
; Sequence 10503, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59, US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
```

```
; SEQ ID NO 10503
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 75
; OTHER INFORMATION: s=g or c
US-09-513-999C-10503

Alignment Scores:
Pred. No.: 1,94e-16 Length: 393
Score: 182.50 Matches: 42
Percent Similarity: 60.23% Conservative: 11
Best Local Similarity: 47.73% Mismatches: 26
Query Match: 19.92% Indels: 10
DB: 4 Gaps: 1

US-10-657-740-1 (1-173) x US-09-513-999C-10503 (1-393)
Qy 81 SerProGluAspLeuThrValHisValGlnAspPheValGluHisGlyHis 100
Db 36 AGCATGAACGAGCTACGCTCAAGACCAAGGATGCGTGGAGATCACCGCAAGCAC 95
Qy 101 AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgTyrArgLeu 120
Db 96 GAGGAGCGCGAGGACGAGCATGGCTACATCTCCCGTCTTCACGCGAAATACACGCTG 155
Qy 121 ProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThr 140
Db 156 CCCCCGGTGGGACCCCAAGTTTCTCTCCCTCCCTGCTCCCTGAGGGGACACTGACC 215
Qy 141 PheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIlePro 160
Db 216 GTGGA-GGCCCC-----CATGCCAAGTAGCCACGCA 247
Qy 161 ValSerArgGluGluLysProThr 168
Db 248 GTCCAACGAGATCACCATCCCACT 271

RESULT 20
US-08-900-407-2
; Sequence 2, Application US/08900407
; Patent No. 5962262
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW HUMAN HEAT SHOCK 27 LIKE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-Seq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,407
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
```

```
; REFERENCE/DOCKET NUMBER: PF-0351 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 1
; IMMEDIATE SOURCE:
; LIBRARY: ????
; CLONE: 1362715
US-08-900-407-2

Alignment Scores:
Pred. No.: 7.19e-14 Length: 1627
Score: 171.50 Matches: 45
Percent Similarity: 47.92% Conservative: 24
Best Local Similarity: 31.25% Mismatches: 58
Query Match: 18.72% Indels: 17
DB: 2 Gaps: 2

US-10-657-740-1 (1-173) x US-08-900-407-2 (1-1627)
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeu----- 36
Db 598 TCTCGCTTGTGGATGATGGCTTTGGCATGAGCCCTTCCAGACGACTTTCACAGCCTCT 657
Qy 37 -----LeuProPheLeuSerSerThrIleSerProTyrTyrArgGlnSer 51
Db 658 TGGCCCGACTGGGCTCTGCTCTGCTCTCTCTCCGCTGCGCCAGGACCCTAAGGTGCGGC 717
Qy 52 LeuPheArgThr-----ValLeuAspSerGlyIleSer 62
Db 718 ATGCTGCCCGCGCGCCCTCCACTGCCACCCAGGTTTGGGGTGGCTGCCAGGCGCAGGACC 777
Qy 63 GluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerPro 82
Db 778 CCCCCACCCTTCCCTGGGAGCCCTGGAAAGTGTGTGATGTGCACAGCTTCAAGCA 837
Qy 83 GluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGlu 102
Db 838 GAGGAGTTGATGTGTGAAGACCAAGATGGATACGTGGAGGTGTCTGGCAACATGAAGAG 897
Qy 103 ArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSer 122
Db 898 AAACAGCAAGAAGGTGGCATTTGTTTCAAGAACTTCAAGAAATAATCCAGCTTCTCTCA 957
Qy 123 AsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCys 142
Db 958 GAGGTGATCTGTGACAGATATTTGCTCTACTTTTCCCGAGAGGGTCTGCTGATCATCGAA 1017
Qy 143 GlyProLysIle 146
Db 1018 GCTCCCCCAGGTCTC 1029

RESULT 21
US-09-513-999C-1048
; Sequence 1048, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duciert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59, US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
```

```

; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1048
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..507
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 51
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 117
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 118
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 159
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 160
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 182
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 183
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 16
; OTHER INFORMATION: Xaa=Asp or Glu
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 38
; OTHER INFORMATION: Xaa=His or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 39
; OTHER INFORMATION: Xaa=Ala or Ser
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 53
; OTHER INFORMATION: Xaa=Glu or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 60
; OTHER INFORMATION: Xaa=Glu or Gly
; OTHER INFORMATION: Xaa=Glu or Gly
US-09-513-999C-1048
Alignment Scores:
Pred. No.: 1..8e-13 Length: 509
Score: 163.00 Matches: 36
Percent Similarity: 68.12% Conservative: 11
Best Local Similarity: 52.17% Mismatches: 16
Query Match: 17.79% Indels: 6
DB: 4 Gaps: 4
US-10-657-740-1 (1-173) x US-09-513-999C-1048 (1-509)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 229 ATGGACATCGCATCCACCACCGCGCCCTTCTTCTTCCCTTCCACTCCGCC 288
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1048
; LENGTH: 509
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4..507
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 51
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 117
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 118
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 159
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 160
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 182
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 183
; OTHER INFORMATION: r=a or g
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 16
; OTHER INFORMATION: Xaa=Asp or Glu
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 38
; OTHER INFORMATION: Xaa=His or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 39
; OTHER INFORMATION: Xaa=Ala or Ser
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 53
; OTHER INFORMATION: Xaa=Glu or Gln
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 60
; OTHER INFORMATION: Xaa=Glu or Gly
; OTHER INFORMATION: Xaa=Glu or Gly
US-09-513-999C-1048
Alignment Scores:
Pred. No.: 1..8e-13 Length: 509
Score: 163.00 Matches: 36
Percent Similarity: 68.12% Conservative: 11
Best Local Similarity: 52.17% Mismatches: 16
Query Match: 17.79% Indels: 6
DB: 4 Gaps: 4
US-10-657-740-1 (1-173) x US-09-513-999C-1048 (1-509)
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 229 ATGGACATCGCATCCACCACCGCGCCCTTCTTCTTCCCTTCCACTCCGCC 288

```

```
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1047
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 284..562
US-09-513-999C-1047

Alignment Scores:
Pred. No.: 2,128-13 Length: 564
Score: 163.00 Matches: 36
Percent Similarity: 68.12% Conservativeness: 11
Best Local Similarity: 52.17% Indels: 16
Query Match: 17.79% Gaps: 4
DB:

US-10-657-740-1 (1-173) x US-09-513-999C-1047 (1-564)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 284 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTTCTTCTTCCACTCCGCC 343
QY 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluGlyAspLeuLeuProPhe 39
Db 344 AGCCGCTCTTGGACAGTCTCTCGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 400
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 401 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTCTCTCGGGCACCAGC 460
QY 56 ValLeuAspSerGlyIleSerGluVal 64
Db 461 TGGTTTGACACTGGAGTCTCAGAGGTG 487

RESULT 24
US-09-513-999C-1049
; Sequence 1049, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1049
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 300..578
US-09-513-999C-1049

Alignment Scores:
Pred. No.: 2,228-13 Length: 580
Score: 163.00 Matches: 36
```

```
Percent Similarity: 68.12% Conservativeness: 11
Best Local Similarity: 52.17% Mismatches: 16
Query Match: 17.79% Indels: 6
DB: 4 Gaps: 4

US-10-657-740-1 (1-173) x US-09-513-999C-1049 (1-580)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 300 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTTCTTCTTCCACTCCGCC 359
QY 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluGlyAspLeuLeuProPhe 39
Db 360 AGCCGCTCTTGGACAGTCTCTCGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 416
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 417 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTCTCTCGGGCACCAGC 476
QY 56 ValLeuAspSerGlyIleSerGluVal 64
Db 477 TGGTTTGACACTGGAGTCTCAGAGGTG 503

RESULT 25
US-09-513-999C-1046
; Sequence 1046, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 1046
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 315..593
US-09-513-999C-1046

Alignment Scores:
Pred. No.: 2,328-13 Length: 595
Score: 163.00 Matches: 36
Percent Similarity: 68.12% Conservativeness: 11
Best Local Similarity: 52.17% Mismatches: 16
Query Match: 17.79% Indels: 6
DB: 4 Gaps: 4

US-10-657-740-1 (1-173) x US-09-513-999C-1046 (1-595)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 315 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTTCTTCTTCCACTCCGCC 374
QY 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluGlyAspLeuLeuProPhe 39
Db 375 AGCCGCTCTTGGACAGTCTCTCGAGAGACCTGTGGAGTCTGATCTTTTCCCG--- 431
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 432 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTCTCTCGGGCACCAGC 491
QY 56 ValLeuAspSerGlyIleSerGluVal 64
Db 432 ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTCTCTCGGGCACCAGC 491
```



```

RESULT 30
US-09-949-016-16734
; Sequence 16734, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03

```

```

RESULT 31
US-09-949-016-89647/c
; Sequence 89647, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89647
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-89647

```



```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 89648
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-89648

Alignment Scores:
Pred. No.:          Length:      6,32e-07             601
Score:              Matches:     21                  21
Percent Similarity: 82.76%       Conservative:    3
Best Local Similarity: 72.41%   Mismatches:     5
Query Match:        Indels:      0                   0
DB:                 Gaps:         0                   0

US-10-657-740-1 (1-173) x US-09-949-016-89648 (1-601)

QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 601 GATGTGAACACTTCTCCCGAGAGAACAATAAAGTTAAAGTGTTGGGAGATGTGATTGAG 542

QY 96 IleHisGLyLysHisAsnGluArgGln 104
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 541 GTGCATGGAAAACATGAGAGGCCCAG 515

RESULT 37
US-09-949-016-17284/c
; Sequence 17284, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ FOR WINDOWS VERSION 4.0
; SEQ ID NO 17284
; LENGTH: 12214
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17284

Alignment Scores:
Pred. No.:          Length:      0.000221           12214
Score:              Matches:     35                  35
Percent Similarity: 51.35%       Conservative:    22
Best Local Similarity: 31.53%   Mismatches:     36
Query Match:        Indels:      18                  18
DB:                 Gaps:         5                   5

US-10-657-740-1 (1-173) x US-09-949-016-17284 (1-12214)

QY 54 ArgThrValLeuAspSerGlyIleSerGluValArgSerAsp-----ArgAspLysPhe 71
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 376 CGGTCCTCAGGACAGTCCAGCGCTCAGGAGCATATGACCATGCCAGAGACGGTTTC 317

QY 72 ValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAsp 91
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 316 CAAATGAAGCTGGATGCCCAACCGCTTCGCCCCGGAGGAACCTGGTGGTGCAGTGGATGGC 257

QY 92 AspPheValGluIleHisGLyLysHisAsn-----GluArgGlnAsp 105
::::::::::: ::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 256 CAATGGCTGATGTCAGCGGACAGCACAGCAACTGGACGCTCAGGAGCCCAGGAGGGTGCT 197

QY 106 AspHisGLyTVrIleSerArgGluPheHisArqArqtVrArqLPurProSerAsnValAsp 125

```

```

Db      393  -----GAGAAGCTGGCGGCTGACGGCACTGTCTATGAACAACCTTCGCTCACAAGTGCCAG 446

Qy      120  LeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 139
        |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  |||||
Db      447  CTGCGGAGGAGCTGGACCCGACGTCGTGTACCTCGGCTCTCTCGGGAGGACGCGAGCCTC 506

Qy      140  Thr 140
        |||
Db      507  ACT 509

RESULT 39
US-09-949-016-17282/c
; Sequence 17282, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17282
; LENGTH: 9615
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-17282

Alignment Scores:
Pred. No.:          0.000778          Length:          9615
Score:             110.00             Matches:         31
Percent Similarity: 51.58%             Conservative:    18
Best Local Similarity: 32.63%           Mismatches:     30
Query Match:       12.01%              Indels:         16
DB:                4                  Gaps:           4

US-10-657-740-1 (1-173) x US-09-949-016-17282 (1-9615)

Qy      68  ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 87
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      9588  AGAGACGGTTTCCAAATGAAGCTGATGCCACGGCTTCGCCCGGAGGAAGCTGGTGGTG 9529

Qy      88  LysValGlnAspAspPheValGluIleHisGlyLysHisAsn----- 101
        |||||  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|  ::::|
Db      9528  CAGGTGGATGGCCAATGGCTGATGGTACCGGACGACGACGACCACTGACGTCAGGACCCG 9469

Qy      102  GluArgGlnAspAspHisGlyTyrlleSerArgGluPheHisArgArgTyArgLeuPro 121
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      9468  GAAAGGGTCAGTTACCGC-----ATGTACAGAAGGTGTCACCGGAAA---ATGCTACCG 9418

Qy      122  SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPhe 141
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      9417  TCCAACTGAGTCTCTACGGCCATGACCTGCTGCCTGACCCCTCGCGGACGCTGGTGTC 9358

Qy      142  CysGly-----ProLysIleGlnThrGly 149
        |||||
Db      9357  AGAGCCCAAGTGTGTGGCGCTGGCCCTCCCTGAAGCCCAACAGGA 9313

RESULT 40
US-09-949-016-14287/c
; Sequence 14287, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

```


; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9018612.3
 ; FILING DATE: 24-AUG-1990
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 861-3000
 ; TELEFAX: (202) 822-0944
 ; TELEX: 6714627 CUSH
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 652 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 US-07-748-761-2

Alignment Scores:

Pred. No.: 0.00228 Length: 652
 Score: 93.50 Matches: 36
 Percent Similarity: 43.9% Conservativity: 26
 Best Local Similarity: 25.53% Mismatches: 56
 Query Match: 10.21% Indels: 23
 DB: 1 Gaps: 7

US-10-657-740-1 (1-173) x US-07-748-761-2 (1-652)

Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu 40
 |||||
 Db 60 AGAATTTTCGGCGATCGAGCAAGCAGCAGCATGTTTCGAT-----CCATTTC 107
 Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
 |||||
 Db 108 ATTGACGTATTGATCCCA---TTCAGGGAATTAGGCTTCCCAAGTACC---AATTTCAGGG 161
 Qy 61 IleSerGluVal-----ArgSerAspArgAspAspVal 72
 |||||
 Db 162 GAGAGCTCTGATTTGCCAACACAGCAATAGATCGAAGAGAACTCCAGAACCTCATGTG 221
 Qy 73 IlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAsp 92
 |||||
 Db 222 TTCAAGGTTGATCTTCCAGGGCTTAAGAAGGAGGAGGAAAGTCAAAGTGAAGTCGAGGAGGAT 281
 Qy 93 PheVal---GluIleHisGlyLysHisAsnGluArgGlnAspAspHisGly----- 108
 |||||
 Db 282 AGGGTTCTTCAGATCAGCGGAGAGAGGAACTGGCAGAGGAAGATGAAGATGAAGTGG 341
 Qy 109 -----TyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsnVal 124
 |||||
 Db 342 CATCGATGAGCGAGCAGCGGGAATTCATGAGGAGATTTAGACTTCCGAGAGATGCA 401
 Qy 125 AspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyPro 144
 |||||
 Db 402 AAGATGGATCAAGTTAAGCGCTCATGGAG---AATGGAGTGTACTGTACTGTTCCTCA 458
 Qy 145 Lys 145
 |||||
 Db 459 AAG 461

RESULT 44

US-09-252-991A-12236
 ; Sequence 12236, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 12236
 ; LENGTH: 465
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-12236

Alignment Scores:

Pred. No.: 0.00216 Length: 465
 Score: 92.00 Matches: 34
 Percent Similarity: 40.58% Conservativity: 22
 Best Local Similarity: 24.64% Mismatches: 50
 Query Match: 10.04% Indels: 32
 DB: 4 Gaps: 5

US-10-657-740-1 (1-173) x US-09-252-991A-12236 (1-465)

Qy 8 ProTrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhePhe 27
 |||||
 Db 40 CCCTGTTCGTCATTCGTTAGGC-----TTCGATCGCTTC--- 75
 Qy 28 GlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIleSerProTyr 47
 |||||
 Db 76 ---AACGATCTGTTGAGTCGCGCTCGCAATGAGCGCGGAGTACTACCCGCCCTAC 132
 Qy 48 TyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAsp 67
 |||||
 Db 133 -----AACGTCGAAAAGCACGCT 150
 Qy 68 ArgAspLysPheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrVal 87
 |||||
 Db 151 GACGACGAGTATCATCGTCGCGCCGCCGCTTCCAGGAAGAGACCTGGACCTG 210
 Qy 88 LysValGlnAspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHis 107
 |||||
 Db 211 CAGTCGAGCGCGGTGTCACCGTCAGTGGCGGCAAGCCGAGAGTCCACGGAACAC 270
 Qy 108 -----GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuPro 121
 |||||
 Db 271 GTGACCTACCTGCACAGGCGCATCGCCAGCGCGCTTCAAGCTGTCTTCGCTCGCC 330
 Qy 122 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 139
 |||||
 Db 331 GATCATATCGAGTCAAGCGGCTTCGCTG-----GCCAAGCGCGCTGCTG 375

RESULT 45

US-09-252-991A-12373/c
 ; Sequence 12373, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 12373
 ; LENGTH: 522
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-12373

Alignment Scores:

Pred. No.: 0.00261 Length: 522
 Score: 92.00 Matches: 34
 Percent Similarity: 40.58% Conservativity: 22
 Best Local Similarity: 24.64% Mismatches: 50
 Query Match: 10.04% Indels: 32

```

DB: 4 Gaps: 5
US-10-657-740-1 (1-173) x US-09-252-991A-12373 (1-522)
Qy 8 ProTrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhePhe 27
Db 429 CCGTGTTCCTCATTCCTCGTAGGC-----TTGATCGCTTC--- 394
Qy 28 GlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeuSerSerThrIleSerProTyr 47
Db 393 ---AACGATCTGTTCGAGTCGGCCTGCGCAATGAGCGCGGAGTAGTACTACCGGCCCTAC 337
Qy 48 TyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAsp 67
Db 336 -----AACGTCGAAAAGCACGGT 319
Qy 68 ArgAspLysPheValIlePheLeuAspValIysHisPheSerProGluAspLeuThrVal 87
Db 318 GACGACGAGTAGTCGATCGTCATCGCCCGCCCGCTTCCAGGAGAAAGACCTGGACCTG 259
Qy 88 LysValGlnAspAspPheValIleHisGlyLysHisAsnGluArgGlnAspAspHis 107
Db 258 CAGTTCGAGCGCGGCTGCTGACCGTCAGTCAGCGGAGCGGAGAGATCCACGAGCAAC 199
Qy 108 -----GlyTyrIleSerArgGluPheHisArgArgTyrArgLeuPro 121
Db 198 GTGACCTACCTGCACGAGGCGATCGCCAGCGCGCTTCAAGTCGTCTCGCCTCGCC 139
Qy 122 SerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeu 139
Db 138 GATCATATCGAGGTCAAGCGCGCTTCGCTG-----GCCAACGGCGCTGCTG 94
RESULT 46
US-09-489-039A-3463
; Sequence 3463, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3463
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3463
Alignment Scores:
Pred. NO.: 0.00306 Length: 576
Score: 92.00 Matches: 30
Percent Similarity: 38.41% Conservative: 28
Best Local Similarity: 19.87% Mismatches: 59
Query Match: 10.04% Indels: 34
DB: 4 Gaps: 4
US-10-657-740-1 (1-173) x US-09-489-039A-3463 (1-576)
Qy 35 AspLeuLeuProPheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeu----- 52
Db 151 GAGCTATTACTATGCGGTAACTTCGATCTTTCCCGCTATATCGTTCAGCCATTGGTTTC 210
Qy 53 -----PheArgThrValLeuAspSerGlyIleSerGluValArgSer----- 66
Db 211 GACCGCTGTTCAACCTACTGGAAAACAATCAAGCCAGACGACGGCGGTACCTCCG 270
Qy 67 -----AspArgAspLysPheValIlePheLeuAspValIysHisPhe 80
Db 271 TATACGTCGAGCTGGTAGACGAAACCACTATCGCATCGGTATCGCGGTGGCTGGCTTT 330

```

```

Db      167288 CTAACCCAGCGCCTACCAAGCGGTGTGTGCGGAGTCCCTCTCTCCCTCTCCCTCCAGAAATG 167222
Qy      166 LysProThr 168
Db      167228 AAGCCACC 167220

RESULT 48
US-09-902-540-6436
; Sequence 6436, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 6436
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-6436

Alignment Scores:
Pred. No.: 0.0169 Length: 489
Score: 86.00 Matches: 33
Percent Similarity: 43.36% Conservative: 16
Best Local Similarity: 29.20% Mismatches: 39
Query Match: 9.39% Indels: 26
DB: 4 Gaps: 4

US-10-657-740-1 (1-173) x US-09-902-540-6436 (1-489)

Qy      63 GluValArgSerAspArgAspLysPheValIlePheLeuAspValLysHisPheSerPro 82
Db      154 GAGGTGAAGGAGACCGAGATGCTTCATCTCAAGCGCGATGTCCTCGGCGGTGGAGAG 213
Qy      83 GluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGlu 102
Db      214 AAGACCTTGAATCACCTCGCGAGAACCGGTGACCATCAGCGCAAGCGGGAAGAA 273
Qy      103 ArgGlnAspAspHisGly-----TyrIleSerArgGlu-----PheHis 115
Db      274 GAGCGCGCGCAGGAAGCGACCGGTACTACGCTACGAGCGCACTATGGCTCGTTTCAGC 333
Qy      116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db      334 CGCACCTTCACGCTCCCGCGGGGTGTG-----AATGCC 366
Qy      136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db      367 GAC-----AACGTCCAGGC-GGACTTCAAGAGTGGCGTCTCT 401
Qy      156 GluArgAla-IleProValSerArgGluGluLysPro 167
Db      402 CAACGTGCGCATCCCAAGAAAGTCCGAGGAGCAGGCC 438

RESULT 49
US-09-902-540-444
; Sequence 444, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

```

; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-..(49487)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11770

Alignment Scores:
Pred. No.: 57 Length: 49487
Score: 84.00 Matches: 46
Percent Similarity: 40.74% Conservatives: 31
Best Local Similarity: 24.34% Mismatches: 54
Query Match: 9.17% Indels: 58
DB: 4 Gaps: 12

US-10-657-740-1 (1-173) x US-09-949-016-11770 (1-49487)

```
QY 34 TyrAspLeuLeuProPheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPhe 53
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2693 TACTACATATGCCATGGGTAAATCA--CTCTCTCGATTCTGTAGGCAGTTCTTTCAC 2637

QY 54 ArgThrValLeuAspSerGlyIleSerGluValArgSerAsp----- 67
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2636 CGC-----TTAGTGAGCTGAGGTGTGAACGACCGACCATAGATAATCA 2595

QY 68 -----ArgAspLysPheValIlePheLeuAsp----- 76
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2594 TCCAGATCGGAAACCCCGAGCGCGCAGGATTATCTGGATAACGGAGCCGGATGGAGGA 2535

QY 77 -----ValLysHisPheSerProGlu-----AspLeuThrValLysVal 89
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2534 TTGTCTCCGTTCTCTCCACACTCAAGCCCGGAAAGGTCACTGACCTAGTCTTATTCTTT 2475

QY 90 GlnAspAspPheValGluIleHisGlyLys---HisAsn----- 101
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2474 ACAGCAGAAATCGGGCAAAACCCATGGCGCCGTCCACAATCTCTATAGCCCGACCTGGG 2415

QY 102 -----GluArgGlnAsp-----AspHisGlyTyrIleSerArg 112
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2414 TACAGGGGAGCAGCGGCGAAACACACACACTTCTAGCGCAGACCGGTTACATTAGGCC 2355

QY 113 GluPheHisArgArgTyrArgLeuProSerAsnVal-----AspGlnSer 127
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2354 GAGCCCAACACACGCGCCACACACACACACACACACACACACACACACACACACACACT 2295

QY 128 AlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGln 147
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2294 TCGTTAACTTGTGATGATAAATTAGGGCCGAGGTGACACTACTGT---CCCGAACTTCAG 2238

QY 148 Thr-----GlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArgGlu 164
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2237 CTGGACGCTAGTCCCAAGTCACTGTCCTAACTGACGCTAGCACTGCCGACCTTGTCCGAG 2178

QY 165 GluLysProThrSerAlaProSerSer 173
   |||  ::  |||:::  |||  :::  |||  |||  |||  |||  |||  |||  |||  |||
Db 2177 GAGTTTCCT---TCCGCTCCCTCAACT 2154
```

Search completed: May 30, 2005, 08:53:36
Job time : 229.078 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 30, 2005, 05:14:41 ; Search time 1216.26 Seconds
(without alignments)
873.287 Million cell updates/sec

Title: US-10-657-740-1

Perfect score: 916

Sequence: 1 MDVTIQHPWFRTKLGPFYPS.....HAERAIPVSRREKTSAPSS 173

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Ygapop 10.0	Ygapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 5700845 seqs, 3069779/57 residues

Total number of hits satisfying chosen parameters: 11401690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10657740/runat_27052005_165254_3380/app_query.fasta_1.590
-DB=Published_Applications NA -QFMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Dblom62
-TRANS=human40.cdi -LIST=150 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=50 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZB=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10657740 @CGN 1.1 904 @runat_27052005_165254_3380
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:**
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	95.1	871	1056	16	US-10-316-253-112 Sequence 112, App
2	89.5	92.7	1271	16	US-10-316-253-110 Sequence 110, App
3	702	76.6	531	13	US-10-105-427-1 Sequence 1, Appli
4	639	69.8	372	18	US-10-657-740-2 Sequence 2, Appli
5	493	53.8	548	15	US-10-101-510-250 Sequence 250, App
6	489	53.4	528	18	US-10-770-668-35 Sequence 35, Appl
7	489	53.4	691	9	US-09-954-456-514 Sequence 514, App
8	489	53.4	691	10	US-09-960-706-869 Sequence 869, App
9	489	53.4	691	10	US-09-873-319-566 Sequence 566, App
10	489	53.4	691	10	US-09-873-367C-764 Sequence 764, App
11	489	53.4	691	17	US-10-133-937-61 Sequence 61, Appl
12	489	53.4	691	17	US-10-172-118-686 Sequence 686, App
13	489	53.4	691	17	US-10-159-563-61 Sequence 61, Appl
14	489	53.4	691	17	US-10-342-887-686 Sequence 686, App
15	489	53.4	691	19	US-10-843-641A-764 Sequence 764, App
16	489	53.4	691	19	US-10-843-641A-3541 Sequence 3541, Ap
17	489	53.4	691	19	US-10-367-057-187 Sequence 187, App
18	485	52.9	528	17	US-10-152-319A-1574 Sequence 1574, Ap
19	485	52.9	689	19	US-10-486-706-455 Sequence 455, App
20	485	52.9	1247	9	US-09-917-800A-1419 Sequence 1419, Ap
21	485	52.9	1247	19	US-10-486-706-105 Sequence 105, App
22	418	45.6	449	9	US-09-960-352-4277 Sequence 4277, Ap
23	394	43.0	396	9	US-09-960-352-12622 Sequence 12622, A
24	367	40.1	211	16	US-10-029-386-26581 Sequence 26581, A
25	367	40.1	573	16	US-10-029-386-26581 Sequence 12881, A
26	364.5	39.8	411	14	US-10-198-846-10360 Sequence 10360, A
27	361	39.4	450	10	US-09-918-995-4949 Sequence 4949, Ap
28	347	37.9	579	16	US-10-029-386-1061 Sequence 1061, Ap
29	347	37.9	834	13	US-10-027-632-172414 Sequence 172414,
30	347	37.9	834	17	US-10-027-632-172414 Sequence 172414,
31	345.5	37.7	829	17	US-10-062-674-1924 Sequence 1924, Ap
32	340	37.1	1820	14	US-10-026-956-320 Sequence 320, App
33	338	36.9	212	16	US-10-723-860-5075 Sequence 5075, Ap
34	338	36.9	421	10	US-10-029-386-14764 Sequence 14764, A
35	338	36.9	421	10	US-09-918-995-4772 Sequence 4772, A
36	336	36.7	486	14	US-10-226-956-303 Sequence 303, App
37	336	36.7	1310	17	US-10-152-319A-1412 Sequence 1412, Ap
38	334	36.5	381	18	US-10-425-115-173881 Sequence 173881,
39	332.5	36.3	380	9	US-09-960-352-4143 Sequence 4143, Ap
40	311.5	34.0	405	10	US-09-918-995-4075 Sequence 4075, Ap
41	296.5	32.4	387	18	US-10-723-860-380 Sequence 380, App
42	295.5	32.3	543	17	US-10-621-901-2279 Sequence 2279, Ap
43	291	31.8	349	9	US-09-900-747-28 Sequence 28, Appl
44	289.5	31.6	764	18	US-10-605-498-91 Sequence 91, Appl
45	289.5	31.6	789	17	US-10-641-643-1300 Sequence 1300, Ap
46	289.5	31.6	847	14	US-10-153-668-283 Sequence 283, App
47	289.5	31.6	865	11	US-09-969-034-4480 Sequence 4480, Ap
48	289.5	31.6	865	17	US-10-172-118-626 Sequence 626, App
49	289.5	31.6	865	17	US-10-342-887-626 Sequence 626, App
50	289.5	31.6	1231	9	US-09-880-107-3865 Sequence 3865, Ap
51	289.5	31.6	1380	14	US-10-100-957A-169 Sequence 169, App
52	289	31.6	787	9	US-09-917-800A-1428 Sequence 1428, Ap
53	288.5	31.5	563	14	US-10-198-846-8376 Sequence 8376, Ap
54	288.5	31.5	787	17	US-10-191-803-228 Sequence 228, App
55	288.5	31.5	787	17	US-10-152-319A-1963 Sequence 1963, Ap
56	288	31.4	599	14	US-10-060-036-43 Sequence 43, Appl
57	287.5	31.4	634	19	US-10-764-420-43 Sequence 43, Appl
58	282.5	30.8	373	17	US-10-242-535A-26398 Sequence 26398, A
59	282.5	30.8	373	17	US-10-085-783A-26398 Sequence 26398, A
60	276	30.1	503	10	US-09-911-904-146 Sequence 146, App
61	266.5	29.1	604	9	US-09-917-800A-981 Sequence 981, App
62	266	29.0	410	18	US-10-468-091-35 Sequence 35, Appl
63	259.5	28.3	654	19	US-10-204-921-16 Sequence 16, Appl
64	257	28.1	1700	9	US-09-813-358-2 Sequence 2, Appli
65	257	28.1	1700	10	US-09-997-279-2 Sequence 2, Appli
66	251	27.4	549	10	US-09-991-936-341 Sequence 341, App
67	250.5	27.3	593	18	US-10-425-115-63403 Sequence 63403, A
68	247	27.0	550	17	US-10-621-901-2194 Sequence 2194, Ap
69	241.5	26.4	708	10	US-09-814-353-14770 Sequence 14770, A

70	238	26.0	874	17	US-10-133-937-62	Sequence 62, Appl	143	110.5	12.1	2152	18	US-10-723-860-395	Sequence 395, App
71	238	26.0	874	17	US-10-159-563-62	Sequence 62, Appl	144	110.5	12.1	2410	17	US-10-104-047-972	Sequence 972, App
72	237	25.9	870	17	US-10-840-038-2	Sequence 3, Appl	145	110.5	12.1	3503	18	US-10-723-860-5087	Sequence 5087, App
73	237	25.9	736	19	US-10-840-038-3	Sequence 3, Appl	146	110	12.0	599	18	US-10-021-323-5328	Sequence 5328, App
74	234.5	25.6	614	10	US-09-814-353-2040	Sequence 2040, App	147	107.5	11.7	462	9	US-09-938-842A-1962	Sequence 1962, App
75	234.5	25.6	614	10	US-09-814-353-3386	Sequence 8386, App	148	107.5	11.7	462	11	US-09-938-842A-1962	Sequence 1962, App
76	233.5	25.5	954	10	US-09-814-353-20273	Sequence 20273, A	149	106.5	11.6	817	18	US-10-767-701-13295	Sequence 13295, A
77	225.5	24.6	542	17	US-10-621-901-2146	Sequence 2146, App	150	106	11.6	3188	18	US-10-425-115-120839	Sequence 120839, App
78	223.5	24.4	500	10	US-09-918-995-22134	Sequence 22134, A							
79	223	24.3	467	14	US-10-060-036-34	Sequence 34, Appl							
80	220.5	24.1	482	10	US-09-918-995-4433	Sequence 4433, App							
c 81	212	23.1	49634	11	US-09-968-007A-514	Sequence 514, App							
c 82	212	23.1	49634	19	US-10-843-641A-6984	Sequence 6984, App							
c 83	205.5	22.4	212	18	US-10-482-158-43	Sequence 43, Appl							
c 84	198.5	21.7	396	9	US-09-954-531-845	Sequence 845, App							
c 85	198.5	21.7	396	9	US-09-954-531-1265	Sequence 1265, App							
c 86	198.5	21.7	396	19	US-10-843-641A-1912	Sequence 1912, App							
c 87	198.5	21.7	396	19	US-10-843-641A-2332	Sequence 2332, App							
c 88	192	21.0	6806	17	US-10-205-331-105	Sequence 105, App							
c 89	191.5	20.9	445	10	US-09-918-995-10509	Sequence 10509, A							
c 90	188	20.5	4206	15	US-10-101-510-288	Sequence 288, App							
c 91	188	20.5	4206	18	US-10-283-975A-93	Sequence 93, Appl							
c 92	187	20.4	2632	13	US-10-044-090-312	Sequence 312, App							
c 93	182.5	19.9	487	17	US-10-369-493-30110	Sequence 30110, A							
c 94	182.5	19.9	487	17	US-10-369-493-30111	Sequence 30111, A							
c 95	180.5	19.7	565	14	US-10-198-846-366	Sequence 366, App							
c 96	174	19.0	510	18	US-10-363-345A-653	Sequence 653, App							
c 97	174	19.0	510	18	US-10-363-345A-654	Sequence 654, App							
c 98	174	19.0	510	19	US-10-363-483A-653	Sequence 653, App							
c 99	174	19.0	510	19	US-10-363-483A-654	Sequence 654, App							
c 100	172.5	18.8	1495	16	US-10-006-285-2629	Sequence 269, App							
c 101	171.5	18.7	1627	14	US-10-116-802-322	Sequence 322, App							
c 102	171.5	18.7	2010	17	US-10-172-118-1482	Sequence 1482, App							
c 103	171.5	18.7	2010	17	US-10-342-887-1482	Sequence 1482, App							
c 104	163.5	17.8	2036	16	US-10-006-285-447	Sequence 447, App							
c 105	159.5	17.4	395	17	US-10-062-674-254	Sequence 254, App							
c 106	159	17.4	395	10	US-09-918-995-4376	Sequence 4376, App							
c 107	158	17.2	810	18	US-10-723-860-4925	Sequence 4925, App							
c 108	149.5	16.3	1891	16	US-10-316-253-108	Sequence 108, App							
c 109	148	16.2	403	9	US-09-960-352-3948	Sequence 3948, App							
c 110	148	16.2	490	17	US-10-369-493-30109	Sequence 30109, A							
c 111	148	16.2	490	17	US-10-369-493-30112	Sequence 30112, A							
c 112	148	16.2	6806	17	US-10-205-331-105	Sequence 105, App							
c 113	146	15.9	442	10	US-09-918-995-5137	Sequence 5137, App							
c 114	142.5	15.6	510	18	US-10-363-345A-655	Sequence 655, App							
c 115	142.5	15.6	510	18	US-10-363-345A-656	Sequence 656, App							
c 116	142.5	15.6	510	19	US-10-363-483A-655	Sequence 655, App							
c 117	142.5	15.6	510	19	US-10-363-483A-656	Sequence 656, App							
c 118	142	15.5	131	16	US-10-029-386-26899	Sequence 26899, A							
c 119	142	15.5	451	15	US-10-102-524-483	Sequence 483, App							
c 120	142	15.5	534	16	US-10-029-386-13199	Sequence 13199, A							
c 121	139	15.2	427	9	US-09-960-352-11415	Sequence 11415, A							
c 122	137.5	15.0	450	9	US-09-880-107-3022	Sequence 3022, App							
c 123	137.5	15.0	450	9	US-09-954-531-962	Sequence 962, App							
c 124	137.5	15.0	450	9	US-09-954-531-1355	Sequence 1355, App							
c 125	137.5	15.0	450	19	US-10-843-641A-2029	Sequence 2029, App							
c 126	137.5	15.0	450	19	US-10-843-641A-2422	Sequence 2422, App							
c 127	131	14.3	303	9	US-09-960-352-2010	Sequence 2010, App							
c 128	128	14.0	446	9	US-09-960-352-5348	Sequence 5348, App							
c 129	127.5	13.9	2153	17	US-10-062-674-2106	Sequence 2106, App							
c 130	125.5	13.7	495	14	US-10-255-536-69	Sequence 69, Appl							
c 131	124	13.5	423	9	US-09-833-790-52	Sequence 52, Appl							
c 132	123	13.4	339	9	US-09-867-701-6118	Sequence 6118, App							
c 133	123	13.4	468	9	US-09-938-842A-2558	Sequence 2558, App							
c 134	123	13.4	468	11	US-09-938-842A-2558	Sequence 2558, App							
c 135	121	13.2	65	10	US-09-908-975-30348	Sequence 30348, A							
c 136	120.5	13.2	42339	13	US-10-087-192-991	Sequence 991, App							
c 137	115.5	12.6	478	10	US-09-918-995-20604	Sequence 20604, A							
c 138	115	12.6	1922	17	US-10-104-047-1686	Sequence 1686, App							
c 139	115	12.6	50002	13	US-10-047-192-994	Sequence 994, App							
c 140	114	12.4	932	18	US-10-425-115-183047	Sequence 183047, App							
c 141	110.5	12.1	719	15	US-10-037-270-669	Sequence 669, App							
c 142	110.5	12.1	719	17	US-10-117-722-669	Sequence 669, App							

ALIGNMENTS

RESULT 1

US-10-316-253-112
; Sequence 112, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M
; CURRENT APPLICATION NUMBER: US/10/316,253
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 112
; LENGTH: 1056
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13)..(534)
; OTHER INFORMATION:
US-10-316-253-112

Alignment Scores:
Pred. No.: 1.42e-108 Length: 1056
Score: 871.00 Matches: 164
Percent Similarity: 97.69% Conservative: 5
Best Local Similarity: 94.80% Mismatches: 4
Query Match: 95.09% Indels: 0
DB: 16 Gaps: 0

US-10-657-740-1 (1-173) x US-10-316-253-112 (1-1056)

Qy	1	MetAspValThrIleGlnHisProTrrPheLysArgThrLeuGlyProPheTyrProSer	20
Db	13	ATGACGTTCACATCCAGCACCCCTTGGTTAAAGCGCGCCCTTACCCACAGC	72
Qy	21	ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu	40
Db	73	CGACTGTTCGACCACTTCTTCGGCGAGGGCCCTTTTGAATACGACCTGCTGCCCTCTCTG	132
Qy	41	SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly	60
Db	133	TCTTCCACCATCAGCCCTACTACCGCCAGTCTCTTCTCCGACAGTGTGGACTCCGGC	192
Qy	61	IleSerGluValArgSerAspArgAspValIlePheLeuAspValIleHisPhe	80
Db	193	ATCTCTGAGGTCCGATCTGACCGGACAAAGTTTGTTCATCTTCTTGGATGTGAAGC	252
Qy	81	SerProGluAspLeuThrValLysValGlnAspAspPheValGluIleHisGlyHis	100
Db	253	TCTCTGAGGACCTCACGTGAGTACTGGAAGATTCGTGGAGATCCATGGCAACAC	312
Qy	101	AsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeu	120
Db	313	AACGAGAGCGGAGGATGACCATGCTACATTTCCCGTGAATTTCCCGCTCGCTACCGTCTG	372

Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db |||||
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db |||||
Qy 156 GluArgAlaIleProValSerArgGluGluLysProThrSerAlaProSerSer 173
Db |||||
Qy 171 ProSerSer 173
Db |||||

RESULT 4

US-10-657-740-2
; Sequence 2, Application US/10657740
; Publication No. US20040157289A1
; GENERAL INFORMATION:
; APPLICANT: Salerno, John C.
; APPLICANT: Hanna, Michael
; APPLICANT: Koretz, Jane F.
; APPLICANT: Crone, Donna
; APPLICANT: Smith, Susan E.
; TITLE OF INVENTION: PROTEIN EXPRESSION SYSTEM
; FILE REFERENCE: 01794100H0406US1
; CURRENT APPLICATION NUMBER: US/10/657,740
; CURRENT FILING DATE: 2003-09-08
; PRIOR APPLICATION NUMBER: US 60/408,680
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-657-740-2

Alignment Scores:
Pred. No.: 1,746-77 Length: 372
Score: 639.00 Matches: 123
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 69.76% Indels: 0
DB: 18 Gaps: 0

US-10-657-740-1 (1-173) x US-10-657-740-2 (1-372)

Qy 51 SerLeuPheArgThrValLeuAspSerGlyIleSerGluValArgSerAspAspLys 70
Db |||||
Qy 71 PheValIlePheLeuAspValLysHisPheSerProGluAspLeuThrValLysValGln 90
Db |||||
Qy 91 AspAspPheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIle 110
Db |||||
Qy 111 SerArgGluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSer 130
Db |||||
Qy 131 CysSerLeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeu 150
Db |||||
Qy 151 AspAlaThrHisAlaGluArgAlaIleProValSerArgGluGluLysProThrSerAla 170
Db |||||
Qy 171 ProSerSer 173
Db |||||

RESULT 5

US-10-101-510-250
; Sequence 250, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 250
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-10-101-510-250

Alignment Scores:
Pred. No.: 3,06e-57 Length: 548
Score: 493.00 Matches: 98
Percent Similarity: 74.16% Conservative: 34
Best Local Similarity: 55.06% Mismatches: 34
Query Match: 53.82% Indels: 12
DB: 15 Gaps: 6

US-10-657-740-1 (1-173) x US-10-101-510-250 (1-548)

Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
Db |||||
Qy 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db |||||
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db |||||
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db |||||
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db |||||
Qy 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db |||||
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db |||||
Qy 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db |||||
Qy 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
Db |||||

RESULT 6

US-10-770-668-35
; Sequence 35, Application US/10770668
; Publication No. US20040191843A1
; GENERAL INFORMATION:

```
; APPLICANT: Wright, Susan C.
; APPLICANT: Larrick, James W.
; APPLICANT: Nock, Steffen R.
; APPLICANT: Wilson, David S.
; TITLE OF INVENTION: Cell-killing Molecules and Methods of Use Thereof
; FILE REFERENCE: ARSALUS-08602
; CURRENT APPLICATION NUMBER: US/10/770,668
; CURRENT FILING DATE: 2004-02-02
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-770-668-35

Alignment Scores:
Pred. No.: 1,02e-56 Length: 528
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 18 Gaps: 6

US-10-657-740-1 (1-173) x US-10-770-668-35 (1-528)
QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 1 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTTCTTCTTCCATCTCCCCC 60
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 61 AGCCGCTCTTTGACAGTCTTCGGAGACACCTGTTGGAGTCTGATCTTTTCCG--- 117
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 118 ACGTCTACTTCCCTGAGTCCCTTACCTTCGGCCACCCTCTCTTCCGCGCACCCAGC 177
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspValIlePheLeu 75
DB 178 TGGTTTGACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 237
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 238 GATGTGAAGCACTTCTCCCGCAGGAACTCAAAGTTAAGGTTTGGAGATGTGATTGAG 297
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 298 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGGAGTTCCAC 357
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 358 AGGAATACCGGATCCCGATGATAGACCTCTCACCATTACTTCTTCCCTGTCATCT 417
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 418 GATGGGTCTCTACGTGATGACCAAGAAACAG-----GTCTCTGGCCCT 465
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 466 GAGCGCACCATTTCCCATCACCCTGAAGAGAGCCCTGTGTGTCACCGCAGCCCC 519

RESULT 7
US-09-954-456-514
; Sequence 514, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
```

```
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 514
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-514

Alignment Scores:
Pred. No.: 1,53e-56 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 9 Gaps: 6

US-10-657-740-1 (1-173) x US-09-954-456-514 (1-691)
QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 26 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTTCTTCTTCCATCTCCCCC 85
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
DB 86 AGCCGCTCTTTGACAGTCTTCGGAGACACCTGTTGGAGTCTGATCTTTTCCG--- 142
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 143 ACGTCTACTTCCCTGAGTCCCTTACCTTCGGCCACCCTCTCTTCCGCGCACCCAGC 202
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspValIlePheLeu 75
DB 203 TGGTTTGACACTGGACTCTCAGAGATGCGCTGGAGAGGACAGGTTCTCTGTCAACCTG 262
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 263 GATGTGAAGCACTTCTCCCGCAGGAACTCAAAGTTAAGGTTTGGAGATGTGATTGAG 322
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 323 GTGCATGGAAACATGAAGAGCGCCAGGATGAACATGGTTTTCATCTCCAGGGAGTTCCAC 382
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 383 AGGAATACCGGATCCCGATGATAGACCTCTCACCATTACTTCTTCCCTGTCATCT 442
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 443 GATGGGTCTCTACGTGATGACCAAGAAACAG-----GTCTCTGGCCCT 490
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 491 GAGCGCACCATTTCCCATCACCCTGAAGAGAGCCCTGTGTGTCACCGCAGCCCC 544
```


US-09-873-367C-764
; Sequence 764, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:

```

: APPLICANT: Young, Paul
: APPLICANT: Soppet, Daniel
: APPLICANT: Endress, Gregory
: APPLICANT: Augustus, Meena
: APPLICANT: Ebner, Reinhard
: APPLICANT: Carter, Kenneth
: TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
: TITLE OF INVENTION: Signature Gene Sets
: FILE REFERENCE: 689290-84
: CURRENT APPLICATION NUMBER: US/09/873,367C
: CURRENT FILING DATE: 2003-04-29
: PRIOR APPLICATION NUMBER: U.S. 60/236,891
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: U.S. 60/236,842
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: U.S. 60/244,867
: PRIOR FILING DATE: 2000-11-01
: PRIOR APPLICATION NUMBER: U.S. 60/245,084
: PRIOR FILING DATE: 2000-11-01
: NUMBER OF SEQ ID NOS: 1067
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 764
: LENGTH: 691
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-873-367C-764

```

Alignment Scores:	
Pred. No.:	1.53e-56
Score:	489.00
Percent Similarity:	73.60%
Best Local Similarity:	54.4%
Query Match:	53.38%
DB:	10
Length:	691
Matches:	97
Conservative:	34
Mismatches:	35
Indels:	12
Gaps:	6

US-10-657-740-1 (1-173) x US-09-873-367C-764 (1-691)

Qy	1	MetAspValThrIleGlnHisProTyrPheIysArgThrLeuGlyProPheTyr---	Pro	19
Db	26	ATGGACATCGCCATCCACACCCCTGGATCGCGCGCCCTCTCTTTCTTCCATCCGCC	85	
Qy	20	SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPhe	39	
Db	86	AGCGCCCTCTTTGACCAGTCTCTCGGAGAGCACCTGTTGGAGTCTGATCTTTTCCCG---	142	
Qy	40	LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr	55	
Db	143	ACGTCTACTTCCCTAGAGTCCCTTCTACCTCTGGGCACCCCTCTCTCTCGCGGCACCCACG	202	
Qy	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu	75	
Db	203	TGGTTTGACACTGGACACTCTAGAGATGCCCTCGGAGAGGACAGGTTCTGTGTCAACCTG	262	
Qy	76	AspValIysHisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGlu	95	
Db	263	GATGTGAAGCACTTCTCCCCAGAGGAATCAAAGTTAAGGTGTTGGGAGATGTGATTGAG	322	
Qy	96	IleHisGlyIysHisAenGluArgGlnAspHisGlyTyrIleSerArgGluPheHis	115	
Db	323	GTGCATGTGAANCAATGAAGAGCGCCAGGATGAACATGGTTTCACTTCCAGGAGGTCCAC	382	
Qy	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla	135	
Db	383	AGGAATACCGGATCCCGAGCTGATGTAGACCCCTCTCACCACTTACTTCTCATCCCTGTCATCT	442	
Qy	136	AspGlyMetLeuThrPheCysGlyProIlysylleGlnThrGlyLeuAspAlaThrHisAla	155	
Db	443	GATGGGGTCTCTCATGTGAATGGACCAAGGAAACAG-----GTCTGTGGCCCT	490	

Qy 156 GluArgAlaIleProValSerArgGluLysPro-----ThrSerAlaPro 179
||||| ||||| :||| :||| :||| :||| :||| :||| :|||
Dδ 491 GAGCGCACATTCCATCACCCTGAAGAAGCCTGTGTCAACGCAGCCCC 540

RESULT 11

```

US-10-133-937-61
; Sequence 61, Application US/10133937
; Publication No. US20030207278A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING,
; TITLE OF INVENTION: DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
; TITLE OF INVENTION: OTHER BIOLOGICAL STATES
; FILE REFERENCE: 11613.56US01
; CURRENT APPLICATION NUMBER: US/10/133,937
; CURRENT FILING DATE: 2002-11-04
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-133-937-61

```

Alignment Scores:	1.53e-56	Length:	691
Pred. No.:	489.00	Matches:	970
Score:	73.60%	Conservative:	343
Percent Similarity:	54.49%	Mismatches:	35
Best Local Similarity:	53.38%	Indels:	12
Query Match:	17	Gaps:	6
DB:			

US-10-657-740-1 (1-173) x US-10-133-937-61 (1-691)

Qy	1	MetaSpValThrIleGlnHisProTrrPhePheLysArgThrLeuGlyProPheTyr---Pro	19
Db	26	ATGGACATCGCCATCCACCAACCCCTGGATCGCGCGCCCTCTTTCTTCCATCCGCC	85
Qy	20	SerArgLeuPheAspGlnPhePheGlyLeuPheGluTyrAspLeuPheProPhe	39
Db	86	AGCGCGCTCTTGACCAAGTCTCTCGGAGAGCACCTGTTGGAGTCGTGATCTTTTC	142
Qy	40	LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr	55
Db	143	ACGTCTACTTCCCTAGTCCCTTCTACTCTCGGCCACCTCTCTTCTGGGCGACCCAGC	202
Qy	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu	75
Db	203	TGGTTTGACATCGACTCTCAGAGATCGCCTGGAGAAGGACAGGTTCTCTGTCAACCTG	262
Qy	76	AspVallyshisPheSerProGluAspLeuThrVallysValGlnAspAspPheValGlu	95
Db	263	GATGTGAAGACATCTCCCGAGAGAACTCAAAGTTAAGGTGTGTGGGAGATGTGATTGAG	322
Qy	96	IleHisGlyLysHisAsnGluAArgGlnAspAspHisGlyTyrIleSerArgGluPheHis	115
Db	323	GTGCATGGAAAACATGAAGAGCGCCAGATGACATGGTTTCATCTCCAGGAGATTCCAC	382
Qy	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla	135
Db	383	AGGAATACCGGATCCCGATCGATGTAGACCTCTCACCATTTACTTTCATCCCTGTCACT	442
Qy	136	AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla	155
Db	443	GATGGGCTCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCT	490
Qy	156	GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro	171
Db	491	GAGCGCACCATTTCCATCAACCGGTGAAGAGAAGCCTGCTGTCCGCGACGCC	544

RESULT 12

US-10-172-118-686
; Sequence 686, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-399
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 686
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM 001895
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-686

Alignment Scores:

Pred. No.: 1.53e-56 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 17 Gaps: 6

US-10-657-740-1 (1-173) x US-10-172-118-686 (1-691)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTCTTTTCCCTTCCACTCCCCC 85
QY 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCCCTTTGACAGTTCTTCGAGAGACACCTGTGGAGTCTGATCTTTTCCCG--- 142
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg---Thr 55
Db 143 ACGTCTACTCCCTGAGTCCCTTCTACCTCGGCCACCTCTCTCTCGGGGACCCAGC 202
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 203 TGGTTTGACACTGGACTCTCAGATGGCCTGGAGAGGACAGGTTCTCTGTCACACCTG 262
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 263 GATGTGAAGCACTTCTCCCAAGAGAACTCAAAGTTAAGGTGTGGGAGATGTGATTGAG 322
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGAAACATGAGAGCCGATGATGACATGTTTCATCTCCAGGGAGTTCCAC 382
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 383 AGGAATACCGATCCCGATCCAGCTGATGATGACCTCTCACCATTACTTCTCATCTGTCACT 442
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGGTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCT 490
QY 156 GluArgAlaIleProValSerArgGluGlyPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTTCCATCACCCTGAGAGAGAGCTGTCTGTACCGCAGCCCCC 544

RESULT 14

Db 491 GAGCGCACCATTTCCATCACCCTGAGAGAGAGCTGTCTGTACCGCAGCCCCC 544

RESULT 13

US-10-159-563-61
; Sequence 61, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613 56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-61

Alignment Scores:

Pred. No.: 1.53e-56 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 17 Gaps: 6

US-10-657-740-1 (1-173) x US-10-159-563-61 (1-691)

QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 26 ATGGACATCGGCATCCACCCCTGGATCGCGCCCTCTTTTCCCTTCCACTCCCCC 85
QY 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluTyrAspLeuLeuProPhe 39
Db 86 AGCCGCCCTTTGACAGTTCTTCGAGAGACACCTGTGGAGTCTGATCTTTTCCCG--- 142
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg---Thr 55
Db 143 ACGTCTACTCCCTGAGTCCCTTCTACCTCGGCCACCTCTCTCTCGGGGACCCAGC 202
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 203 TGGTTTGACACTGGACTCTCAGATGGCCTGGAGAGGACAGGTTCTCTGTCACACCTG 262
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 263 GATGTGAAGCACTTCTCCCAAGAGAACTCAAAGTTAAGGTGTGGGAGATGTGATTGAG 322
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 323 GTGCATGAAACATGAGAGCCGATGATGACCTCTCACCATTACTTCTCATCTGTCACT 382
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
Db 383 AGGAATACCGATCCCGATCCAGCTGATGATGACCTCTCACCATTACTTCTCATCTGTCACT 442
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
Db 443 GATGGGGTCTCTCACTGTGAATGGACCAAGAAACAG-----GTCTCTGGCCCT 490
QY 156 GluArgAlaIleProValSerArgGluGlyPro-----ThrSerAlaPro 171
Db 491 GAGCGCACCATTTCCATCACCCTGAGAGAGAGCTGTCTGTACCGCAGCCCCC 544

US-10-342-887-686
; Sequence 686, Application US/10342887
; Publication No. US20040059340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 686
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-686

Alignment Scores:
Pred. No.: 1.53e-56 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 17 Gaps: 6

US-10-657-740-1 (1-173) x US-10-342-887-686 (1-691)

QY	1	MetAspValThrIleGlnHisProTyrPheLeuArgThrLeuGlyProPheTyr---	Pro 19
DB	26	ATGACATCCCATCCCATCCACCCCTGGATCGCGCCCTCTTTCTTTCCACTCCCC	85
QY	20	SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuPhe	39
DB	86	AGCGCCCTTTTTCACCATTTCTTCGAGAGACCTGTGTGAGTCTGATCTTTCCCG	142
QY	40	LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg----- <td>Thr 55</td>	Thr 55
DB	143	ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTTCCTCGGGGACCC	202
QY	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe	Leu 75
DB	203	TGGTTTGACATCGACTCTCAGAGATCGCTCGAGAGACAGGTCTCTGTCAACCTG	262
QY	76	AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspAspPheVal	Glu 95
DB	263	GATGTGAAGCACTTCTCCCGAGAGAACTCAAGATTAAAGTGTGGGAGATGTGATTG	322
QY	96	IleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe	His 115
DB	323	GTGCATGGAAAAATGAAGACGCCAGGATGAACATGTTTCACTCCAGGGAGTTCCAC	382
QY	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer	Ala 135
DB	383	AGAAAATACCGATCCAGCTGATGATGAGCCCTCTCACCATTCTTCCTCTGTCATCT	442
QY	136	AspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAla	155
DB	443	GATGGGGTCTCTACTGTGAATGATGACCAAGGAACAG-----GTCTCTGCCCT	490
QY	156	GluArgAlaIleProValSerArgGluGluLeuPro-----ThrSerAlaPro	171

US-10-657-740-1 (1-173) x US-10-843-641A-764 (1-691)

QY	1	MetAspValThrIleGlnHisProTyrPheLeuArgThrLeuGlyProPheTyr---	Pro 19
DB	26	ATGACATCCCATCCCATCCACCCCTGGATCGCGCCCTCTTTCTTTCCACTCCCC	85
QY	20	SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuPhe	39
DB	86	AGCGCCCTTTTTCACCATTTCTTCGAGAGACCTGTGTGAGTCTGATCTTTCCCG	142
QY	40	LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg----- <td>Thr 55</td>	Thr 55
DB	143	ACGTCTACTTCCCTGAGTCCCTTCTACCTTCGGCCACCCCTCTTCCTCGGGGACCC	202
QY	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe	Leu 75
DB	203	TGGTTTGACATCGACTCTCAGAGATCGCTCGAGAGACAGGTCTCTGTCAACCTG	262
QY	76	AspValIleHisPheSerProGluAspLeuThrValIleValGlnAspAspPheVal	Glu 95
DB	263	GATGTGAAGCACTTCTCCCGAGAGAACTCAAGATTAAAGTGTGGGAGATGTGATTG	322
QY	96	IleHisGlyIleHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe	His 115
DB	323	GTGCATGGAAAAATGAAGACGCCAGGATGAACATGTTTCACTCCAGGGAGTTCCAC	382
QY	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer	Ala 135
DB	383	AGAAAATACCGATCCAGCTGATGATGAGCCCTCTCACCATTCTTCCTCTGTCATCT	442
QY	136	AspGlyMetLeuThrPheCysGlyProIleGlnThrGlyLeuAspAlaThrHisAla	155
DB	443	GATGGGGTCTCTACTGTGAATGATGACCAAGGAACAG-----GTCTCTGCCCT	490
QY	156	GluArgAlaIleProValSerArgGluGluLeuPro-----ThrSerAlaPro	171

US-10-843-641A-764
; Sequence 764, Application US/10843641A
; Publication No. US20050064454A1
; GENERAL INFORMATION:
; APPLICANT: Avalon Pharmaceuticals, Inc.
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; TITLE OF INVENTION: Signature Gene Sets
; FILE REFERENCE: 689290-189
; CURRENT APPLICATION NUMBER: US/10/843,641A
; CURRENT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: US/09/873,367
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/09/954,456
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,436
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/962,832
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/09/964,824
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US/09/967,768
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/09/968,007
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,347
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/09/969,708
; PRIOR FILING DATE: 2001-10-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 8447
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 764
; LENGTH: 691
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-843-641A-764

Alignment Scores:
Pred. No.: 1.53e-56 Length: 691
Score: 489.00 Matches: 97
Percent Similarity: 73.60% Conservative: 34
Best Local Similarity: 54.49% Mismatches: 35
Query Match: 53.38% Indels: 12
DB: 19 Gaps: 6

Qy	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgAspGlyAspPheValIlePheLeu	75
Db	203	TGGTTTCACACTGGACTCTCAGAGATGGCCCTGGAGAAGGACAGGTTCTCTGTCACACCTG	262
Qy	76	AspValIlyeHisPheSerProGluAspLeuThrValIlyeValGlnAspAspPheValGlu	95
Db	263	GATGTGAAGCACTTCTCCCAAGAGAACTCAAAAGTTAAAGGTTCGGGAGATGTAATTGAG	322
Qy	96	IleHisGlyLyHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis	115
Db	323	GTGCATGGANAACATGAAGCGCCAGGATGAACATGGTTTCATCTCCAGGAGATTCAC	382
Qy	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla	135
Db	383	AGGAAATACCGGATCCCGACTGATGTAGACCTCTCACCATTACTTCCTCTGTCATCT	442
Qy	136	AspGlyMetLeuThrPheCysGlyProIlyleGlnThrGlyLeuAspAlaThrHisAla	155
Db	443	GATGGGGTCCCTCACTGTGAATGGACCAAGGAACAG-----GTCTCTGGCCCT	490
Qy	156	GluArgAlaIleProValSerArgGluGluIysPro-----ThrSerAlaPro	171
Db	491	GAGGCGACCAATCCCATCACCCTGAAGAGAAGCCTGTGTCCAGCGACCCCC	544

RESULT 18

US-10-152-319A-1574
; Sequence 1574, Application US/10152319A
; Publication No. US20040072160A1

```

; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling

```

FILE REFERENCE: 44921-5089-US
CURRENT APPLICATION NUMBER: US/10/152,333
CURRENT FILING DATE: 2002-05-22

; PRIOR APPLICATION NUMBER: US 60/292,335
 ;
 ; PRIOR FILING DATE: 2001-05-22
 ;
 ; PRIOR APPLICATION NUMBER: US 60/297,523

; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: US 60/298,925
 ; PRIOR FILING DATE: 2001-06-19

; PRIOR FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US 60/303,810
 ; PRIOR FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US 60/303,807

; PRIOR APPLICATION NUMBER: US 60/303,808
 ; PRIOR FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US 60/303,808
 ; PRIOR FILING DATE: 2001-07-10

; PRIOR FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: US 60/315,047
 ;
 ; PRIOR FILING DATE: 2001-08-28
 ; PRIOR APPLICATION NUMBER: US 60/324,928

; PRIOR APPLICATION NUMBER: US 60/324,324
 ; PRIOR FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/330,867
 ; PRIOR FILING DATE: 2001-11-01

; PRIOR FILING DATE: 2001-11-01
 ; PRIOR APPLICATION NUMBER: US 60/330,462
 ; PRIOR FILING DATE: 2001-10-22
 ; Remaining Prior Application data removed

```

; REMAINING PRIOR APPLICATION DATA REMOVE
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1574

```

Percent Similarity:	74.16%	Conservative:	36
Best Local Similarity:	53.93%	Mismatches:	34
Query Match:	52.95%	Indels:	12
DB:	17	Gaps:	6

US-10-657-740-1 (1-173) x US-10-152-319A-1574 (1-528)

Qy	1	MetAspValThrIleGlnHisProTrrPheIysArgThrLeuGlyProPheTyr---	Pro	19
Db	1	ATGGACATGCCATCCACCAACCCCTGGATCGGGGTCCCTTCTTTCTTCCATCTCCCA	60	
Qy	20	SerArgLeuPheAspGlnPhePheGlyGluClyLeuPheGluTyrAspLeuLeuProPhe	39	
Db	61	AGCGCGCTCTTGACCAGTTCTTCGGAGAGCACCTGTGGAGTCTGACCTCTTCTCT	117	
Qy	40	LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr	55	
Db	118	ACAGCCACTCCCTGAGCCCTTCTACTTCGGCCACCCTCTCTCTCGGGGCACCTAGC	177	
Qy	56	ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu	75	
Db	178	TGGATTGACACTGGGCTCTCAGAGATCGGTATGGAGAAGGACAGGTTCTCTGTGAACCTG	237	
Qy	76	AspValIysHisPheSerProGluAspLeuThrValIysValGlnAspAspPheValGlu	95	
Db	238	GACGTGAAGCACTTCTCCAGAGGAACCTCAAAGTCAAGGTTCTCGGGAGACGTGATTCAG	297	
Qy	96	IleHisGlyIysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis	115	
Db	298	GTGCACGGCAAGCACGAAGAAGCCCGACGAGCAACATGGCTTCATCTCCAGGGAGTCCAC	357	
Qy	116	ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla	135	
Db	358	AGGAAGTACCGGATCCAGCCGGAGCTGGATCTCTCACCATTACTTCTTCCCTGTCTATCG	417	
Qy	136	AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla	155	
Db	418	GATGGAGTCTCACTGTGAATGGACCAAGGAACAG-----GCCTCTGGCCCT	465	
Qy	156	GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro	171	
Db	466	GAGCGCACCATTCCTCATCCGCTGAAGAAGAAGCTGCTGTCATCTGAGGCCCT	519	

RESULT 19

US-10-486-706-455
; Sequence 455, Application US/10486706
; Publication No. US20050071088A1

; PUBLICATION NO. US20030071088A1
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: LANDFIELD, PHILIP W.
 ;
 ; APPLICANT: BLALOCK, ERIC M.

; AFFILIANT: BLALOCK, ERIC M.
 ; APPLICANT: CHEN, KUEY-CHU
 ; APPLICANT: FOSTER, THOMAS C.
 ; TITLE OF INVENTION: GENE EXPRES

```

; TITLE OF INVENTION: BRAIN AGING AND AGE-
; TITLE OF INVENTION: GENE EXPRESSION PROF-
; TITLE OF INVENTION: GENE EXPRESSION PROF-
; FILE REFERENCE: 50229-426
; CURRENT APPLICATION NUMBER: IIS/10/486.706

```

; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/US02/25607
; CURRENT FILING DATE: 2002-08-13

; PRIOR FILING DATE: 2002-08-13
 ; PRIOR APPLICATION NUMBER: US 60/311,343
 ; PRIOR FILING DATE: 2001-08-13
 ; NUMBER OF SEQ ID NOS: 451

```

; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 455
; LENGTH: 598

```

```

; LENGTH: 689
; TYPE: DNA
; ORGANISM: Rattus rattus
ITS 10.486.706.455

```

Alignment Scores:	
Pred. No.:	5,37e-56
Score:	485.00
Percent Similarity:	74.16%
Conservative:	36
Length:	689
Matches:	96

Best Local Similarity: 53.93% Mismatches: 34
Query Match: 52.95% Indels: 12
DB: 19 Gaps: 6

US-10-657-740-1 (1-173) x US-10-486-706-455 (1-689)

```
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 13 ATGGACATAGCCATCCACACCCCTGGATCGGCGTCCCTCTTCTTCCCTTCCACTCCCA 72
QY 20 SerArgLeuPheAspGlnPheGlyLeuGlyLeuPheGluTyrAspLeuProPhe 39
DB 73 AGCCGCTCTTTGACAGTCTTCGAGACACCTGTTGGAGTCGTGACCTCTTTCT--- 129
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 130 ACAGCCTCTTCCCTGAGCCCTTCTACCTTCGGCCACCTCTCTCTCGGGACCTTAGC 189
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 190 TGGATTGACATGGGCTCTCAGATGCGTATGGAGAGGACAGGTTCTCTGTGACCTG 249
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 250 GACGTGAAGCACTTCTCCAGAGCACTCAAAGTCAAGGTTCTCGGAGACGTGATTGAG 309
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 310 GTGCACGGCAAGACAGAGCGCCAGGACGACGACGACGACGACGACGACGACGACG 369
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 370 AGGAATGACCGATCCCGACGCGGACGATGATCTCTCACCATTACTTCTCTCTGTCATCG 429
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 430 GATGAGTCTCTCAGTGTGAATGGACCAAGGAACAG-----GCCTCTGGCCCT 477
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 478 GAGCGCACCATTCCTCCATCACCCTGAGAGAGCGCTGCTGTCTACTGCGACCCCT 531
```

RESULT 20

US-09-917-800A-1419
; Sequence 1419, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; PRIORITY FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19

; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1419
; LENGTH: 1247
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 M55534
; US-09-917-800A-1419

Alignment Scores:

Pred. No.:	1,31e-55	Length:	1247
Score:	485.00	Matches:	96
Percent Similarity:	74.16%	Conservative:	36
Best Local Similarity:	53.93%	Mismatches:	34
Query Match:	52.95%	Indels:	12
DB:	9	Gaps:	6

US-10-657-740-1 (1-173) x US-09-917-800A-1419 (1-1247)

```
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 583 ATGGACATAGCCATCCACACCCCTGGATCGGCGTCCCTCTTCTTCCCTTCCACTCCCA 642
QY 20 SerArgLeuPheAspGlnPheGlyLeuGlyLeuPheGluTyrAspLeuProPhe 39
DB 643 AGCCGCTCTTTGACAGTCTTCTCGGAGACACCTGTTGGAGTCGTGACCTCTTCTCT--- 699
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 700 ACAGCCTCTTCCCTGAGCCCTTCTACCTTCGGCCACCTCTCTCTCGGGACCTTAGC 759
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 760 TGGATTGACATGGGCTCTCAGATGCGTATGGAGAGGACAGGTTCTCTGTGAACTCG 819
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 820 GACGTGAAGCACTTCTCTCCAGAGCACTCAAAGTCAAGGTTCTCGGAGACGTGATTGAG 879
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 880 GTGCACGGCAAGACAGAGCGCCAGGACGACGACGACGACGACGACGACGACGACG 939
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 940 AGGAATGACCGATCCCGACGCGGACGCGGACGCGGACGCGGACGCGGACGCGGACG 999
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 1000 GATGAGTCTCTCAGTGTGAATGGACCAAGGAACAG-----GCCTCTGGCCCT 1047
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 1048 GAGCGCACCATTCCTCCATCACCCTGAGAGAGCGCTGCTGTCTACTGCGACCCCT 1101
```

RESULT 21

US-10-486-706-105
; Sequence 105, Application US/10486706
; Publication No. US20050071088A1
; GENERAL INFORMATION:
; APPLICANT: LANDFIELD, PHILIP W.
; APPLICANT: BLALOCK, ERIC M.
; APPLICANT: CHEN, KUEY-CHU
; APPLICANT: FOSTER, THOMAS C.
; TITLE OF INVENTION: GENE EXPRESSION PROFILE BIOMARKERS AND THERAPEUTIC TARGETS FOR
; FILE REFERENCE: 50229-426
; CURRENT APPLICATION NUMBER: US/10/486,706
; CURRENT FILING DATE: 2004-02-13
; PRIOR APPLICATION NUMBER: PCT/US02/25607

;; PRIOR FILING DATE: 2002-08-13
;; PRIOR APPLICATION NUMBER: US 60/311,343
;; PRIOR FILING DATE: 2001-08-13
;; NUMBER OF SEQ ID NOS: 461
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 105
;; LENGTH: 1247
;; TYPE: DNA
;; ORGANISM: Rattus norvegicus
US-10-486-706-105

Alignment Scores:
Pred. No.: 1,318-55 Length: 1247
Score: 485.00 Matches: 96
Percent Similarity: 74.1% Conservative: 36
Best Local Similarity: 53.9% Mismatches: 34
Query Match: 52.9% Indels: 12
DB: 19 Gaps: 6

US-10-657-740-1 (1-173) x US-10-486-706-105 (1-1247)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 593 ATGGACATAGCCATCCACCACCTGGATCGGCGTCCCTTCTTCTTCCACTCCCA 642
QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
DB 643 AGCGCCCTCTTTGACAGTCTTCGGAGAGACCTGTTGGAGTCTGACCTCTTCTCT--- 699
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 700 ACAGCCATCTCCCTGAGCCCTTCTACCTTCGGCCACCTCTCTCTCGGCGCACCTAGC 759
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 760 TGGATTGACCTGGGCTCTCAGAGATGCTGATGGAGAGACAGATTCTCTGTGAACCTG 819
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 820 GACGTGAAGCACTTCTCTCCAGAGAACTCAAGTCAAGTTCGGAGACGTGATTGAG 879
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 880 GTGCACGGCAAGCAGCAAGAGCGCCAGGACCAATGGCTTCATCTCCAGGGAGTTCCAC 939
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 940 AGGAAGTACCGGATCCAGCGGACCGGATGATCTCTCACCATTACTTCTCTGTCATCG 999
QY 136 AspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAla 155
DB 1000 GATGAGTCTCTCCTGATGATGACCAAGAAACAG-----GCCTCTGCCCT 1047
QY 156 GluArgAlaIleProValSerArgGluGluLysPro-----ThrSerAlaPro 171
DB 1048 GAGCGCACCATTCCCATCCCGTGAAGAGAGCGCTGTGTCTCAGTCCAGCCCT 1101

RESULT 22

US-09-960-352-4277
; Sequence 4277, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; PRIOR FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4277

;; LENGTH: 449
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; OTHER INFORMATION: Clone ID: 19-BOWMS1-021-Q1-E1-E3
US-09-960-352-4277

Alignment Scores:
Pred. No.: 4,22e-47 Length: 449
Score: 418.00 Matches: 79
Percent Similarity: 77.30% Conservative: 30
Best Local Similarity: 56.03% Mismatches: 26
Query Match: 45.63% Indels: 6
DB: 9 Gaps: 4

US-10-657-740-1 (1-173) x US-09-960-352-4277 (1-449)

QY 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 30 ATGGATATCGCATCCACCACCTGGATCGGCGCTTCTTCTTCTTCCACTCTCCC 89
QY 20 SerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
DB 90 AGCGCCCTCTTTGACAGTCTTTTGGCGAGCACCTGTTAGAGTCTGATCTCTTCCA--- 146
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 147 GCTTCTACTCTCCCTGAGCCCTTCTACCTTCGGCGCCCTCATTTCTGCGGCACCCAGC 206
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 207 TGGATTGACCTGGGCTCTCAGAGATGCTGATGGAGAGACAGATTCTCTGTCAACCTG 266
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 267 GATGTGAAGCACTTCTCTCCAGAGAACTCAAGTCAAGTTCGGAGATGTGATTGAG 326
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 327 GTGCATGCAACATGAAGAGCGCCAGGATGAACATGGTTTTATCTCCCGGAGGTCCAC 386
QY 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAla 135
DB 387 AGGAATACCGGATCCAGCTGACGTGGAACCTCTCGCCATTACTTCTCTCTGCTCT 446
QY 136 Asp 136
DB 447 GAT 449

RESULT 23

US-09-960-352-12622
; Sequence 12622, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12622
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 54-LIB3057-007-Q1-K1-F6
US-09-960-352-12622
Alignment Scores:
Pred. No.: 6,75e-44 Length: 396
Score: 394.00 Matches: 74

Percent Similarity: 77.52% Conservative: 26
Best Local Similarity: 57.36% Mismatches: 23
Query Match: 43.01% Indels: 6
DB: 9 Gaps: 4

US-10-657-740-1 (1-173) x US-09-960-352-12622 (1-396)

```
QY 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
DB 13 ATGGATATCGCCATCACACCCCTGGATCCGCCGCCCTTCTTCCCTTTCCACTCTCCC 72
QY 20 SerArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuProPhe 39
DB 73 AGCCGCTCTTGACCACTTTTTCGGCAGCACCTGTGGAGTCTGTCTTCCA--- 129
QY 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
DB 130 GCTTCTACTTCCCTGAGGCCCTTCTACCTTCGCCGCCCTCATTTCTCGGCGCACCCAGC 189
QY 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
DB 190 TGGATTCACACTGGCCCTCTCAGAGATCGCTTGGAGAGGACAGATTCTCTGTCAACCTG 249
QY 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
DB 250 GATGTGAAGCACTTCTCCCGAGGAACTCAAGGTCAAGTCTCGGAGATGTGATTGAG 309
QY 96 IleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
DB 310 GTGCATGGCAACATGAAAGAGCGCAGATGAATGTTTATCTCCGGGAGTTCCAC 369
QY 116 ArgArgTyrArgLeuProSerAsnVal 124
DB 370 AGGAATACCGGATCCCGACTGACGTG 396
```

RESULT 24

```
US-10-029-386-26581
; Sequence 26581, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 26581
; LENGTH: 211
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: SWISSPROT HIT: P02489, EVALUATE 3.00e-36
; OTHER INFORMATION: EST HUMAN HIT: BF726399.1, EVALUATE 1.00e-113
; OTHER INFORMATION: NT HIT: g114780619, EVALUATE 1.00e-115
```

Alignment Scores:

Pred. No.:	1.31e-40	Length:	211
Score:	367.00	Matches:	70
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.07%	Indels:	0
DB:	16	Gaps:	0

```
US-10-657-740-1 (1-173) x US-10-029-386-26581 (1-211)
QY 104 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 123
DB 1 CAGGACGACCAACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCGCTCCAAAC 60
QY 124 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 143
DB 61 GTGGACCACTGGCCCTCTTGTCTCCCTGTCTGCCGATGGCATGCTGACCTTCTGTGGC 120
QY 144 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 163
DB 121 CCAAGATCCAGACTGGCTGGATGCCACCCAGCGAGCGAGCCATCCCGGTGCGCG 180
QY 164 GluGluLysProThrSerAlaProSerSer 173
DB 181 GAGGAGAAGCCACCTCGGCTCCCTCGTCC 210
```

RESULT 25

```
US-10-029-386-12881
; Sequence 12881, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12881
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.8
; OTHER INFORMATION: NT HIT: AF026952.1, EVALUATE 0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: BF726856.1, EVALUATE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P02489, EVALUATE 1.00e-32
```

Alignment Scores:

Pred. No.:	5.86e-40	Length:	573
Score:	367.00	Matches:	70
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	40.07%	Indels:	0
DB:	16	Gaps:	0

US-10-657-740-1 (1-173) x US-10-029-386-12881 (1-573)

```
QY 104 GlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSerAsn 123
DB 203 CAGGACGACCAACCGCTACATTTCCCGTGAGTTCCACCGCGCTACCGCTGCGCTCCAAAC 262
QY 124 ValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCysGly 143
DB 263 GTGGACCACTGGCCCTCTTGTCTCCCTGTCTGCCGATGGCATCTGACCTTCTGTGGC 322
QY 144 ProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAlaIleProValSerArg 163
DB 323 CCAAGATCCAGACTGGCTGGATGCCACCCAGCGAGCGAGCCATCCCGGTGCGCG 382
QY 164 GluGluLysProThrSerAlaProSerSer 173
DB 383 GAGGAGAAGCCACCTCGGCTCCCTCGTCC 412
```

```
RESULT 26
US-10-198-846-10360
; Sequence 10360, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10360
; LENGTH: 411
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-10360

Alignment Scores:
Pred. No.: 7,82e-40 Length: 411
Score: 364.50 Matches: 71
Percent Similarity: 75.40% Conservative: 24
Best Local Similarity: 56.35% Mismatches: 24
Query Match: 39.79% Indels: 7
DB: 14 Gaps: 5

US-10-657-740-1 (1-173) x US-10-198-846-10360 (1-411)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 39 ATGGACATAGCATCCACACCCCTGGATCGCGGTCTCTTTTCCCTTTCCACTCCCA 98
Qy 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluTyrAspLeuProPhe 39
Db 99 AGCGCCCTCTTTGACCACTTCTCGAGAGACCTGTGGAGTCTGACCTCTCTCTCT 155
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGln-----SerLeuPheArg-----Thr 55
Db 156 ACAGCCACTCTCCCTGAGCCCTCTTCTACCTTCGCCACCCCTCTCTCTCGCGGACCTAGC 215
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 216 TGGATTGACACTGGGCTCTCAGAGATGCGTATGGAGAGACAGGTTCTCTGTGAACCTG 275
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 276 GACGTGAAGCACTTCTCTCCAGAGAACTCAAAGTCAAGGTTCTGGGAGAGCGTATTGAG 335
Qy 96 IleHisGlyLysHisAsnGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 336 GTGCACGGCAGCAGGAGAGCGCCAGGACGACGACGACGACGACGACGACGACGACG 395

RESULT 27
US-09-918-995-4949
; Sequence 4949, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
```

```
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4949
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1)---(450)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-4949

Alignment Scores:
Pred. No.: 2.7e-39 Length: 450
Score: 361.00 Matches: 73
Percent Similarity: 71.94% Conservative: 27
Best Local Similarity: 52.52% Mismatches: 33
Query Match: 39.41% Indels: 6
DB: 10 Gaps: 4

US-10-657-740-1 (1-173) x US-09-918-995-4949 (1-450)

Qy 1 MetAspValThrIleGlnHisProTrpPheLysArgThrLeuGlyProPheTyr---Pro 19
Db 37 ATGGACATCGCATCCACCGCCCTGGATCGCGACCCCTTTTGTGCTTTCCACTCCCCC 96
Qy 20 SerArgLeuPheAspGlnPheGluGlyLeuPheGluTyrAspLeuProPhe 39
Db 97 AGCGCCCTCTTTGACCACTTCTCGAGAGACCTGTGGAGTCTGATCTTTGCCCG--- 153
Qy 40 LeuSerSerThrIleSerProTyrTyrArgGlnSer-----LeuPheArg-----Thr 55
Db 154 ACGTCTACTTCCCTGAGTCCCTTCTACTCTCGGTACCCCTCTTACTCGGACCCAGC 213
Qy 56 ValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeu 75
Db 214 TGGTGTGACACTGACACTCTCAAAGATCGCTGAGAGAGACAGGCTCTGTGTCAACCTG 273
Qy 76 AspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheValGlu 95
Db 274 GATGTGAAGCACTTCTCTCCAGAGAACTCAGAGTTAATGTGTGTGATGAGATTGAG 333
Qy 96 IleHisGlyLysHisAsnGlnAspAspHisGlyTyrIleSerArgGluPheHis 115
Db 334 GTGCATGGAAGCACTGAGAGCGCCAGATGAACTGTTTCTATCTCCAGGAGTTCCAC 393
Qy 116 ArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
Db 394 AGGACATACCGATCCCGCTGATGATAGACCCCTCTGACCATTTACTTTCATCCCTGTCN 450

RESULT 28
US-10-029-386-1061
; Sequence 1061, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; TITLE OF INVENTION: HANZEL, DAVID K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1061
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Homo sapiens
```

FEATURE:
; OTHER INFORMATION: MAP TO CHR21 69.0
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: P02488, EVALUE 5.00e-33
; OTHER INFORMATION: NT HIT: g114780622, EVALUE 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BF726222.1, EVALUE 0.00e+00
US-10-029-386-1061

Alignment Scores:
Pred. No.: 3.26e-37 Length: 579
Score: 347.00 Matches: 65
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.88% Indels: 0
DB: 16 Gaps: 0

US-10-657-740-1 (1-173) x US-10-029-386-1061 (1-579)

Qy 1 MetaSpValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTy-ProSer 20
Db 146 ATGGAGTGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCGAGC 205
Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 206 CGGCTGTTCCAGCAGTATTTTCGGGAGGGCCCTTTTGGAGTATGACCTGCTGCCCTTCCTG 265
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 266 TCGTCCACCATCAGCCCTACTACCGCAGTCCCTCTTCCGACCGTGTGACTCCGGC 325
Qy 61 IleSerGluValArg 65
Db 326 ATCTCTGAGGTAGA 340

RESULT 29
US-10-632-172414
; Sequence 172414, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 172414
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-172414

Alignment Scores:
Pred. No.: 3.26e-37 Length: 579
Score: 347.00 Matches: 65
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.88% Indels: 0
DB: 16 Gaps: 0

Alignment Scores:
Pred. No.: 5.64e-37 Length: 834
Score: 347.00 Matches: 65
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.88% Indels: 0
DB: 13 Gaps: 0

US-10-657-740-1 (1-173) x US-10-027-632-172414 (1-834)

Qy 1 MetaSpValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTy-ProSer 20
Db 251 ATGGAGTGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCGAGC 310
Qy 21 ArgLeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 311 CGGCTGTTCCAGCAGTATTTTCGGGAGGGCCCTTTTGGAGTATGACCTGCTGCCCTTCCTG 370
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 371 TCGTCCACCATCAGCCCTACTACCGCAGTCCCTCTTCCGACCGTGTGACTCCGGC 430
Qy 61 IleSerGluValArg 65
Db 431 ATCTCTGAGGTAGA 445

RESULT 30

US-10-027-632-172414
; Sequence 172414, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 172414
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-172414

Alignment Scores:
Pred. No.: 5.64e-37 Length: 834
Score: 347.00 Matches: 65
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 37.88% Indels: 0
DB: 17 Gaps: 0

US-10-657-740-1 (1-173) x US-10-027-632-172414 (1-834)

Qy 1 MetaSpValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTy-ProSer 20
Db 251 ATGGAGTGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCGAGC 310


```
Db 424 CTCAGTATC-----CAAGCA 438
Qy 159 IleProValSerArgGluGluValProThrSerAla 170
Db 439 GCCCCCGCGTCCAGCCAGCCCGCTCCGCGTCT 474

RESULT 33
US-10-723-860-5075
; Sequence 5075, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5075
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-5075

Alignment Scores:
Pred. No.: 1,66e-35 Length: 1820
Score: 340.00 Matches: 69
Percent Similarity: 58.14% Conservative: 31
Best Local Similarity: 40.12% Mismatches: 50
Query Match: 37.12% Indels: 22
DB: 18 Gaps: 3

US-10-657-740-1 (1-173) x US-10-723-860-5075 (1-1820)
Qy 3 ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPhe-----Tyr 18
Db 35 GTGCTGTGCGAGCGCTCTTGGCTGCGCGCGCTCGCGCGCTTCCCGGACTTTCGCGC 94
Qy 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
Db 95 CCGGACGCGCTCTTTGACGAGCGCTTCGCGGAGGGGCTGCTGGAGCGCGAGCTGCGT 154
Qy 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
Db 155 CTCTGCCCGCCACGCTCGCCCTCTACTAC-----CTGCGGCGACCCAGCGTGGCG 205
Qy 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVallys 78
Db 206 CTGCGCGTCCGCGAGTCCGAGCGACCCCGCGCCACTTTTCGGTGTCTGTAGAGTGAAG 265
Qy 79 HisPheSerProGluAspLeuThrVallysValGlnAspPheValGluIleHisGly 98
Db 266 CACTTCTCCCGGAGGAAATGCTGTCAAGGTGTGGCGGCAACACGCTGGAGGTGCACGCG 325
Qy 99 LysHisAsnGluArgGlnAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118
Db 326 CGCCACGAGGAGCGCCGAGATGAGCGGATTCGTCGCGCGGAGTTCCACCGTCCCTAC 385
Qy 119 ArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMet 138
Db 386 CGCTGCGCGCTGGCGTGGATCCGCTCGCGTACGCTGCTGCTGCTGCTGCTGCTGCTG 445
Qy 139 LeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHisAlaGluArgAla 158
Db 446 CTGTCCATC-----CAGGCC 460
Qy 159 IleProValSerArgGluGluValProThrSerAla 170
```

```
Db 461 GCACCAGCGTCGCCAGGCCAGGCCAGGCCAGGCCAGGCCAG 496
RESULT 34
US-10-029-386-14764
; Sequence 14764, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14764
; LENGTH: 212
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21 69.0
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: NT HIT: U05569.1, EVALUE 1.00e-116
; OTHER INFORMATION: EST_HUMAN HIT: BF727348.1, EVALUE 1.00e-116
; OTHER INFORMATION: SWISSPROT HIT: P02493, EVALUE 9.00e-33
US-10-029-386-14764

Alignment Scores:
Pred. No.: 1,23e-36 Length: 212
Score: 338.00 Matches: 63
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 36.90% Indels: 0
DB: 16 Gaps: 0

US-10-657-740-1 (1-173) x US-10-029-386-14764 (1-212)
Qy 1 MetAspValThrIleGlnHisProTyrPheLysArgThrLeuGlyProPheTyrProSer 20
Db 24 ATGGACGTGACCATCCAGCACCCCTGGTTCAAGCGCACCCCTGGGGCCCTTCTACCCCGC 83
Qy 21 ArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu 40
Db 84 CGGCTGTTCGACCACTTTTCGCGGAGGCGCTTTTGTAGTATGACCTCTGCCCCCTCTCTG 143
Qy 41 SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGly 60
Db 144 TCGTCCACCATCAGCCCTTACTACGCGAGTCCCTCTTCGCGACCGTCTGGACTCCGCG 203
Qy 61 IleSerGlu 63
Db 204 ATCTCTGAG 212

RESULT 35
US-09-918-995-4772
; Sequence 4772, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
```


Pred. No.: 1.67e-35 Length: 380
Score: 332.50 Matches: 65
Percent Similarity: 75.65% Conservative: 22
Best Local Similarity: 56.52% Mismatches: 23
Query Match: 36.30% Indels: 5
DB: 9 Gaps: 2

US-10-657-740-1 (1-173) x US-09-960-352-4143 (1-380)

QY 54 ArgThrValLeuAsp---SerGlyIleSerGluValArgSerAspArgAspLysPheVal 72
DB 12 CGCACCAGTGATGCTCTGCCCCCTCAGAGATGGCTGGAGAGACAGATTCTCT 71
QY 73 IlePheLeuAspVallyHisPheSerProGluAspLeuThrVallyValGlnAspAsp 92
DB 72 GTCAACCTGGACCTGACGACCTCTCCCGACAGGAACCTCAAGGCCAAGGTGCTGGAGAT 131
QY 93 PheValGluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArg 112
DB 132 GTGATTGAGGTGATGGCAACATGAAGAGCGCCAGATGAACATGTTTATCTCCCG 191
QY 113 GluPheHisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSer 132
DB 192 GAGTCCACAGGAATACCGATCCAGCTGACGTGGACCTCTGCCATTACTTCATCC 251
QY 133 LeuSerAlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAla 152
DB 252 CTGTCTCTGATGGGACCTCACTGTGAATGGACCAAGGAACAG-----GCC 299
QY 153 ThrHisAlaGluArgAlaIleProValSerArgGluGluLysPro 167
DB 300 TCCGGCCCTGAGCGCACCATCCCATACCCCTGGAAGAGACCG 344

RESULT 40

US-09-918-995-4075
; Sequence 4075, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4075
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-4075

Alignment Scores:
Pred. No.: 1.38e-32 Length: 405
Score: 311.50 Matches: 60
Percent Similarity: 63.64% Conservative: 24
Best Local Similarity: 45.45% Mismatches: 41
Query Match: 34.01% Indels: 7
DB: 10 Gaps: 2

US-10-657-740-1 (1-173) x US-09-918-995-4075 (1-405)

QY 3 ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProphe-----Tyr 18
DB 16 GTGCTGTGACGCGCTTTGGTGGCGCGCGCTCGGCCCTTCCCGAGACTTTTCGGCG 75
QY 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
DB 76 CCGGACGCGCTTTTGACACGCGCTTCGCGGAGGGGCTGCTGGAGCGCGAGCTGGCTGG 135
QY 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58

DB 136 CTCTGCCCCACCAAGCTCGCCCCCTACTAC-----CTGCGCGCACCCACGCTGGCG 186
QY 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVally 78
DB 187 CTGCGCGTGCAGGTCGCGAGACCCCGGCCACTTTTCGGTGTCTAGACGTGAAG 246
QY 79 HisPheSerProGluAspLeuThrVallyValGlnAspAspPheValGluIleHisGly 98
DB 247 CACTTCTCGCGGAGGAATTCGTCAAGGTGGTGGCGAACACACGTCGAGGTGCACGCG 306
QY 99 LysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118
DB 307 CCCCACGAGGACGCGCGGATGAGCAGGATTCGTGCGCGCGAGTTCACCGCTCGCTAC 366
QY 119 ArgLeuProSerAsnValAspGlnSerAlaLeuSer 130
DB 367 CGCCTGCCCGCTGGCGTGGATCCGCTGCCGTGACG 402

RESULT 41

US-10-723-860-380
; Sequence 380, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnick, Wendy M.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 380
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-380

Alignment Scores:
Pred. No.: 1.46e-30 Length: 387
Score: 296.50 Matches: 57
Percent Similarity: 63.41% Conservative: 21
Best Local Similarity: 46.34% Mismatches: 38
Query Match: 32.37% Indels: 7
DB: 18 Gaps: 2

US-10-657-740-1 (1-173) x US-10-723-860-380 (1-387)

QY 3 ValThrIleGlnHisProTyrPheLysArgThrLeuGlyProphe-----Tyr 18
DB 25 GTGCTGTGACGCGCTTTGGTGGCGCGCGCTCGGCCCTTCCCGAGACTTTTCGGCG 84
QY 19 ProSerArgLeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuPro 38
DB 85 CCGGACGCGCTTTTGACACGCGCTTCGCGGAGGGGCTGCTGGAGCGCGAGCTGGCTGG 144
QY 39 PheLeuSerSerThrIleSerProTyrTyrArgGlnSerLeuPheArgThrValLeuAsp 58
DB 145 CTCTGCCCCACCAAGCTCGCCCCCTACTAC-----CTGCGCGCACCCACGCTGGCG 195
QY 59 SerGlyIleSerGluValArgSerAspArgAspLysPheValIlePheLeuAspVally 78
DB 196 CTGCGCGTGCAGGTCGCGAGACCCCGGCCACTTTTCGGTGTCTAGACGTGAAG 255
QY 79 HisPheSerProGluAspLeuThrVallyValGlnAspAspPheValGluIleHisGly 98
DB 256 CACTTCTCGCGGAGGAATTCGTCAAGGTGGTGGCGGAACACGTCGAGGTGCACGCG 315
QY 99 LysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyr 118

Db 316 CCGCAGGAGCGCCGCGATGACGCGATTCGTCGCGCGAGTTCACCGTCGCTAC 375
Qy 119 ArgLeuPro 121
Db 376 CGCCTGCGC 384

RESULT 42

US-10-621-901-2279
; Sequence 2279, Application US/10621901
; Publication No. US20040067516A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE NUCLEIC ACID
; FILE REFERENCE: PC-8-C3
; CURRENT APPLICATION NUMBER: US/10/621,901
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/319,414
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 2313
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2279
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-10-621-901-2279

Alignment Scores:
Pred. No.: 3,34e-30 Length: 543
Score: 295.50 Matches: 64
Percent Similarity: 60.14% Conservative: 22
Best Local Similarity: 44.76% Mismatches: 50
Query Match: 32.26% Indels: 7
DB: 17 Gaps: 2

US-10-657-740-1 (1-173) x US-10-621-901-2279 (1-543)

Qy 7 HisProTrpPheLysArgThrLeuGlyProPheTyrProSerArgLeuPheAspGlnPhe 26
Db 120 CAGGACTGGTGGAG-----GATTACGACCAACCAATGCTTATGGACCAACAT 170
Qy 27 PheGlyGluClyLeuPheGluTyrAspLeuPheProPheLeuSerSerThrIleSer--- 45
Db 171 TTTGGAATGGCCCTAAATCGTAAATGATCTTATTACTTAATCTAAGGGCCCACTCCATCACTC 230
Qy 46 -----ProTyrTyrArgGlnSerLeuPheArgThrValLeuAspSerGlyIleSer 62
Db 231 TTCGGTGGCGGTTATTACAGACCTTGGAGGATGAATTAATGCTGACGATTCCTTCAUCA 290
Qy 63 GluValArgSerAspArgLysPheValIlePheLeuAspValLysHisPheSerPro 82
Db 291 ACTATCGTCTGACAAAGATAATTCGAAGTCACTTTAGACGTTCAACAATTCAAACCA 350
Qy 83 GluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLysHisAsnGlu 102
Db 351 AAAGAAATCACCCTGAAAAACAAGAGCAATTTGCGTAATCGTCGAAGGCAACACGAGAA 410
Qy 103 ArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArgLeuProSer 122
Db 411 AAACAGATGACCAACGATATATTTCCCGTCATTTTGTTCGAGGATATGTTTGGCCAGA 470
Qy 123 AsnValAspGlnSerAlaLeuSerCysSerLeuSerAlaAspGlyMetLeuThrPheCys 142
Db 471 AATCAGATGCTCCCGATGATGATCGAGTTTGTCTCGATGAGTATTTGACCAATAACA 530
Qy 143 GlyProLys 145
Db 531 GCGCCCAAG 539

RESULT 43

US-09-990-747-28
; Sequence 28, Application US/09990747
; Publication No. US20020081688A1
; GENERAL INFORMATION:
; APPLICANT: Kamb et al.
; TITLE OF INVENTION: Retinoid Pathway Assays, and Compositions Therefrom
; FILE REFERENCE: 29345/36934A
; CURRENT APPLICATION NUMBER: US/09/990,747
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 60/249,468
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 08/812,994
; PRIOR FILING DATE: 1997-03-04
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
; LENGTH: 349
; TYPE: DNA
; ORGANISM: Vaccinia virus
US-09-990-747-28

Alignment Scores:
Pred. No.: 7.1e-30 Length: 349
Score: 291.00 Matches: 57
Percent Similarity: 76.42% Conservative: 24
Best Local Similarity: 53.77% Mismatches: 19
Query Match: 31.77% Indels: 6
DB: 9 Gaps: 4

US-10-657-740-1 (1-173) x US-09-990-747-28 (1-349)

Qy 24 AspGlnPhePheGlyGluClyLeuPheGluTyrAspLeuPheProPheLeuSerThr 43
Db 24 GACGTCTACTTCCCTGAGTCCGTG---GAGTCTGATCTTTCCCG---ACGCTACTTCC 77
Qy 44 IleSerProTyrTyrArgGln-----SerLeuPheArg-----ThrValLeuAspSer 59
Db 78 CTGAGTCCCTCTACTCTTCGGCCACCTCTCTCTCGGGCACCAGCTGGTTTGACACT 137
Qy 60 GlyIleSerGluValArgSerAspArgLysPheValIlePheLeuAspValLysHis 79
Db 138 GCGACTCTCAGAGATCGCTGGAGAGGACAGGTTCTCTCAACTGGATGTGAAGCAC 197
Qy 80 PheSerProGluAspLeuThrValLysValGlnAspPheValGluIleHisGlyLys 99
Db 198 TTCTCCCGCAGAGAACTCAAGTTAAGGTGTGGAGATGTGATGGAGTGCATGGAAGA 257
Qy 100 HisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPheHisArgArgTyrArg 119
Db 258 CATGAAGAGCGCCAGGATGAACATGTTTCATCTCCAGGGAGTTCACAGGAATACCGG 317
Qy 120 LeuProSerAsnValAsp 125
Db 318 ATCCAGCTGATGTAGAC 335

RESULT 44

US-10-605-498-91
; Sequence 91, Application US/10605498
; Publication No. US20040127441A1
; GENERAL INFORMATION:
; APPLICANT: Gleave, Martin
; APPLICANT: Rocchi, Palma
; APPLICANT: Signaevsky, Maxim
; TITLE OF INVENTION: Compositions and Methods for Treatment of Prostate and Other
; FILE REFERENCE: UBC.P-031
; CURRENT APPLICATION NUMBER: US/10/605,498
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: US 60/415,859
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: US 60/463,952
; PRIOR FILING DATE: 2003-04-18

; NUMBER OF SEQ ID NOS: 91
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 91
 ; LENGTH: 764
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-605-498-91

Alignment Scores:
 Pred. No.: 3,7e-29 Length: 764
 Score: 289.50 Matches: 75
 Percent Similarity: 50.52% Conservatives: 23
 Best Local Similarity: 38.66% Mismatches: 43
 Query Match: 31.60% Indels: 54
 DB: 18 Gaps: 6

US-10-657-740-1 (1-173) x US-10-605-498-91 (1-764)

```

QY 10 PheLysArgThrLeuGlyPro-----PheTyrPro---SerArg 21
Db 47 TTCTCGCTCTCGCGGCGCCAGCTGGGACCCCTTCCGCGACTGGTACCGCATAGCCGC 106
QY 22 LeuPheAspGlnPheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu--- 40
Db 107 CTCTTCGACCAAGCCCTTCGGG-----CTGCCCGCGGTGCGG 142
QY 40 ----- 40
Db 143 GAGGAGTGTGCGAGCTGGTTAGCGCGCAGCAGCTGCGCAGGCTACGTGCGCCCTGCCC 202
QY 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
Db 203 CCGCGCGCCATCGAGAGCCCGCAGTGGCGCGCGCCCTACAGCGCGCGCTCAGCCGG 262
QY 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspAspAspLysPheValIlePhe 74
Db 263 CAA----CTCAGCAGCGGGTCTCGGAGATCCGCGCACACTGCGGACCGCTGGCGCGTGC 319
QY 75 LeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheVal 94
Db 320 CTGGATGTCAACACTTCCCGCGGACGAGCTGACGTCAGACCAAGATGGCGGTGGTG 379
QY 95 GluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe 114
Db 380 GAGATCACCAGCAAGCAGCAGGAGCGGCGGACGAGCATGGTACATCTCCCGGTGCTTC 439
QY 115 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
Db 440 ACGCGAATATACAGCTGCCCGCGGTGGAGCCCGCCACCAAGTTTCTCTCTCCCTGTC 499
QY 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
Db 500 CTGAGGGGCACACTGACCGTGA-GGCCCC-----CAT 531.

QY 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
Db 532 GCCCAAGTAGCCAGCAGTCCACAGGATCACCATCCAGT 573
  
```

RESULT 45

US-10-641-643-1300
 ; Sequence 1300, Application US/10641643
 ; Publication No. US20040077003A1
 ; GENERAL INFORMATION:

; APPLICANT: Cocks, Benjamin G.
 ; Susan G. Stuart
 ; Jeffrey J. Seilhamer
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
 ; GENE EXPRESSION

; NUMBER OF SEQUENCES: 1508
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 ; STREET: 3174 PORTER DRIVE
 ; CITY: PALO ALTO

; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/641,643
 ; FILING DATE: 14-Aug-2003
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: <Unknown>
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Zeller, Karen J.
 ; REGISTRATION NUMBER: 37,071
 ; REFERENCE/DOCKET NUMBER: PA-0001 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (650) 855-0555
 ; TELEFAX: (650) 845-4166

INFORMATION FOR SEQ ID NO: 1300:

SEQUENCE CHARACTERISTICS:
 LENGTH: 789 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GENBANK
 CLONE: g32477
 SEQUENCE DESCRIPTION: SEQ ID NO: 1300 :
 US-10-641-643-1300

Alignment Scores:
 Pred. No.: 3,88e-29 Length: 789
 Score: 289.50 Matches: 75
 Percent Similarity: 50.52% Conservatives: 23
 Best Local Similarity: 38.66% Mismatches: 43
 Query Match: 31.60% Indels: 54
 DB: 17 Gaps: 6

US-10-657-740-1 (1-173) x US-10-641-643-1300 (1-789)

```

QY 10 PheLysArgThrLeuGlyPro-----PheTyrPro---SerArg 21
Db 74 TTCTCGCTCTCGCGGCGCCAGCTGGGACCCCTTCCGCGACTGGTACCGCATAGCCGC 133
QY 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuLeuProPheLeu--- 40
Db 134 CTCTTCGACCAAGCCCTTCGGG-----CTGCCCGCGGTGCGG 169
QY 40 ----- 40
Db 170 GAGGAGTGTGCGAGTGTAGCGCGCAGCAGCTGCGCAGGCTACGTGCGCCCTGCCC 229
QY 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
Db 230 CCGCGCGCATCGAGAGCCCGCGCGCTCAGCGCGCGCTCAGCCGG 289
QY 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspAspAspLysPheValIlePhe 74
Db 290 CAA----CTCAGCAGCGGGTCTCGGAGATCCGCGCACACTGCGGACCGCTGGCGGTGTC 346
QY 75 LeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheVal 94
Db 347 CTGGATGTCAACACTTCCCGCGCGAGCTGACGTCAGCGTCAAGACCAAGATGGCGGTG 406
QY 95 GluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe 114
Db 407 GAGATCACCAGCAAGCAGGAGCGGCGGACGAGCATGGTGTATCTCTCCCGGTGCTTC 466
QY 115 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
  
```

```
Db 467 ACCGCGAATACACAGCTGCGCCCGGTGGAGCCACCAAGTTTCTCTCTCCCTGTC 526
      |||::||| ||||| ||||| ::|::| |||||
Qy 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
      |||::||| ||||| ||||| |||||
Db 527 CTGAGGGGACACCTGACCGTGA-GGCCCC-
      |||::||| ||||| ||||| |||||
Qy 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
      |||::||| ||||| ||||| |||||
Db 559 GCCCAAGTAGCCAGCAGCTCCACGAGATCACCATCCAGT 600
      |||::||| ||||| ||||| |||||

RESULT 46
US-10-153-668-283
; Sequence 283, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 283
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (108)..(722)
US-10-153-668-283

Alignment Scores:
Pred. No.: 4,32e-29 Length: 847
Score: 289.50 Matches: 75
Percent Similarity: 50.52% Conservative: 23
Best Local Similarity: 38.66% Mismatches: 43
Query Match: 31.60% Indels: 54
DB: 14 Gaps: 6

US-10-657-740-1 (1-173) x US-10-153-668-283 (1-847)
Qy 10 PheLysArgThrLeuGlyPro-----PheTyrPro---SerArg 21
      |||::||| ||||| ||||| |||||
Db 129 TTCTCGCTCTCGCGGGGCCAGCTGGGACCCCTTCGCGAGCTGGTACCCGCATAGCCGC 188
      |||::||| ||||| ||||| |||||
Qy 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu--- 40
      |||::||| ||||| ||||| |||||
Db 189 CTCTTCGACAGCCCTTCGGG-----CTGCCCCGGCTGCCG 224
      |||::||| ||||| ||||| |||||
Qy 40 ----- 40
Db 225 GAGGAGTGTGTCAGTGGTTAGCGGCAGCAGCTGAGGCTACGTGGCCCCCTGCC 284
      |||::||| ||||| ||||| |||||
Qy 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
      |||::||| ||||| ||||| |||||
Db 285 CCGCGCGCCATCGAGAGCCCGCAGTGGCGCGCGCCCTACAGCGCGCGCTCAGCCGG 344
      |||::||| ||||| ||||| |||||
```

```
Qy 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 74
      |||::||| ||||| ||||| |||||
Db 345 CAA---CTCAGCAGCGGGTCTCGGAGATCCGGACACATGCGGACCGCTGGCGGTGCC 401
      |||::||| ||||| ||||| |||||
Qy 75 LeuAspValLysHisPheSerProGluAspLeuThrValLysValGlnAspAspPheVal 94
      |||::||| ||||| ||||| |||||
Db 402 CTGGATGTCAACCACTTCGCCCGGACGAGCTGACGGTCAAGACCAAGATGCGCGTGGTG 461
      |||::||| ||||| ||||| |||||
Qy 95 GluIleHisGlyLysHisAsnGluArgGlnAspAspHisGlyTyrIleSerArgGluPhe 114
      |||::||| ||||| ||||| |||||
Db 462 GAGATCACCAGGACGAGGAGCGGAGCAGCAGCATGGCTATCTCCCGGTGCTTC 521
      |||::||| ||||| ||||| |||||
Qy 115 HisArgArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
      |||::||| ||||| ||||| |||||
Db 522 ACGCGAAATACACGCTGCCCGCGTGTGGACCCACCCCAAGTTTCTCTCTCCCTGCC 581
      |||::||| ||||| ||||| |||||
Qy 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
      |||::||| ||||| ||||| |||||
Db 582 CCGTGGGGCACACTGACCGTGA-GGCCCC-
      |||::||| ||||| ||||| |||||
Qy 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
      |||::||| ||||| ||||| |||||
Db 614 GCCCAAGTAGCCAGCAGCTCCACGAGATCACCATCCAGT 655
      |||::||| ||||| ||||| |||||

RESULT 47
US-09-969-034-4480
; Sequence 4480, Application US/09969034
; Publication No. US20040110668A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Astle, Jon H.
; APPLICANT: Carroll, Eddie III
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dwivedi, Poornima
; APPLICANT: Molino, Gary A.
; APPLICANT: Thiagalingam, Arunthathi
; APPLICANT: Lewis, Marcia E.
; TITLE OF INVENTION: Nucleic Acid Sequences Differentially
; FILE REFERENCE: 1657/1032
; CURRENT APPLICATION NUMBER: US/09/969,034
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/237,271
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 4494
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4480
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-034-4480

Alignment Scores:
Pred. No.: 4.45e-29 Length: 865
Score: 289.50 Matches: 75
Percent Similarity: 50.52% Conservative: 23
Best Local Similarity: 38.66% Mismatches: 43
Query Match: 31.60% Indels: 54
DB: 11 Gaps: 6

US-10-657-740-1 (1-173) x US-09-969-034-4480 (1-865)
Qy 10 PheLysArgThrLeuGlyPro-----PheTyrPro---SerArg 21
      |||::||| ||||| ||||| |||||
Db 129 TTCTCGCTCTCGCGGGGCCAGCTGGGACCCCTTCGCGAGCTGGTACCCGCATAGCCGC 188
      |||::||| ||||| ||||| |||||
Qy 22 LeuPheAspGlnPhePheGlyGluGlyLeuPheGluTyrAspLeuProPheLeu--- 40
      |||::||| ||||| ||||| |||||
Db 189 CTCTTCGACAGCCCTTCGGG-----CTGCCCCGGCTGCCG 224
      |||::||| ||||| ||||| |||||
Qy 40 ----- 40
Db 225 GAGGAGTGTGTCAGTGGTTAGCGGCAGCAGCTGAGGCTACGTGGCCCCCTGCC 284
      |||::||| ||||| ||||| |||||
```

```

QY 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 CCGCGCCCATCGAGAGCCCGCAGTGGCGCGCGCCCTACAGCCGCGCTCAGCCGG 344
QY 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 74
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 345 CAA---CTCAGCAGCGGGGTCTCGAGATCCGGACACACTGCGGACCGCTGCGGTGC 401
QY 75 LeuAspValIleHisPheSerProGluAspLeuThrValIysValGlnAspPheVal 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 CTGGATGTCAACACATTCGCCCGGACGAGCTGACCGTCAAGACCAAGATGCGGTG 461
QY 95 GluIleHisGlyLysHisGlnArgGlnAspPheGlyTyrIleSerArgGluPhe 114
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 GAGATCACCGGCAAGCAGGAGCGGACGAGCATGGCTACATCTCCCGGTGCTTC 521
QY 115 HisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 ACGCGAAATACACGCTGCCCGCGGTGTGGACCCCAAGTTTCTCTCCCTGCTCC 581
QY 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 CTGAGGGCACACTGACCGTGA-GGCCCC-----CAT 613
QY 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 614 GCCCAAGCTAGCCAGCAGTCCACGAGATCACCATCCCACT 655

```

RESULT 48

```

US-10-172-118-626
; Sequence 626, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 626
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_001540
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-626

```

Alignment Scores:

```

Pred. No.: 4,45e-29 Length: 865
Score: 289.50 Matches: 75
Percent Similarity: 50.52% Conservative: 23
Best Local Similarity: 38.66% Mismatches: 43
Query Match: 31.60% Indels: 54
DB: 17 Gaps: 6

```

US-10-657-740-1 (1-173) x US-10-172-118-626 (1-865)

```

QY 10 PheLysArgThrLeuLeuPro-----PheTyrPro---SerArg 21
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 129 TTCTCGCTCCTCGGGGCCCCAGCTGGAGACCCCTTCGGGATGGTACCCGATGCCGC 188
QY 22 LeuPheAspGlnPhePheGlyGluLeuPheGluTyrAspLeuLeuProPheLeu--- 40

```

```

Db 189 CTCCTTCAGCAGCCCTTCGGG-----CTGCCCCCGCTGCCG 224
QY 40 -----
Db 225 GAGGAGTGTGTCAGTGGTTAGCGGCGCAGCAGTGGCCAGGCTACGTGGCCCCCTGCC 284
QY 41 -----SerSerThrIleSerProTyrTyrArgGlnSerLeuPheArg 54
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 CCGCGCCCATCGAGAGCCCGCAGTGGCGCGCGCCCTACAGCCGCGCTCAGCCGG 344
QY 55 ThrValLeuAspSerGlyIleSerGluValArgSerAspArgAspLysPheValIlePhe 74
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 345 CAA---CTCAGCAGCGGGGTCTCGAGATCCGGACACACTGCGGACCGCTGCGGTGC 401
QY 75 LeuAspValIleHisPheSerProGluAspLeuThrValIysValGlnAspPheVal 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 CTGGATGTCAACACATTCGCCCGGACGAGCTGACCGTCAAGACCAAGATGCGGTG 461
QY 95 GluIleHisGlyLysHisGlnArgGlnAspPheGlyTyrIleSerArgGluPhe 114
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 GAGATCACCGGCAAGCAGGAGCGGACGAGCATGGCTACATCTCCCGGTGCTTC 521
QY 115 HisArgTyrArgLeuProSerAsnValAspGlnSerAlaLeuSerCysSerLeuSer 134
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 ACGCGAAATACACGCTGCCCGCGGTGTGGACCCCAAGTTTCTCTCCCTGCTCC 581
QY 135 AlaAspGlyMetLeuThrPheCysGlyProLysIleGlnThrGlyLeuAspAlaThrHis 154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 CTGAGGGCACACTGACCGTGA-GGCCCC-----CAT 613
QY 155 AlaGluArgAlaIleProValSerArgGluGluLysProThr 168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 614 GCCCAAGCTAGCCAGCAGTCCACGAGATCACCATCCCACT 655

```

RESULT 49

```

US-10-342-887-626
; Sequence 626, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 626
; LENGTH: 865
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-626

```

Alignment Scores:

```

Pred. No.: 4,45e-29 Length: 865
Score: 289.50 Matches: 75
Percent Similarity: 50.52% Conservative: 23
Best Local Similarity: 38.66% Mismatches: 43
Query Match: 31.60% Indels: 54
DB: 17 Gaps: 6

```

